**UNIVERSIDAD AUTÓNOMA DE MADRID**
**DEPARTMENT OF MOLECULAR BIOLOGY**
FACULTY OF SCIENCES

# Structural and proteomic analysis of the human MCM2-7/GINS/Cdc45 assembly

Jaime Martinez Gago
TESIS DOCTORAL
Madrid, 2014

**UNIVERSIDAD AUTÓNOMA DE MADRID**
**DEPARTMENT OF MOLECULAR BIOLOGY**
FACULTY OF SCIENCES

# Structural and proteomic analysis of the human MCM2-7/GINS/Cdc45 assembly

**Jaime Martínez Gago**

Thesis Directors:
**Dr. Guillermo Montoya Blanco**
Centro Nacional de Investigaciones Oncológicas (CNIO)
**Dr. Jasminka Bošković**
Centro Nacional de Investigaciones Oncológicas (CNIO)

Madrid, June 2014

**Dr. Guillermo Montoya Blanco**, jefe del grupo de Cristalografía de Macromoléculas del Centro Nacional de Investigaciones Oncológicas.

**Dr. Jasminka Bošković Vrančić**, investigadora del grupo de Cristalografía de Macromoléculas del Centro Nacional de Investigaciones Oncológicas.

**Dr. Jose María Valpuesta Moralejo**, jefe del Departamento de Estructura de Macromoléculas del Centro Nacional de Biotecnología

Certifica que:

Jaime Martinez Gago, licenciado en Biología y Genética, ha realizado bajo nuestra dirección el trabajo "hMCM2-7 helicases", en el Centro Nacional de Investigaciones Oncológicas.

Consideramos satisfactorio el trabajo realizado y apto para ser presentado como Tesis Doctoral.

Y para que conste a los efectos oportunos, firmo el preente Certificado en Madrid, a – de Septiembre de 2014

| Director | Co-director | Tutor |
|---|---|---|
| Guillermo Montoya | Jasminka Bošković | Jose María Valpuesta |

**This Thesis was supported by a 'Formación de personal investigador' (FPI) fellowship (Ministerio de Educación y Ciencia) awarded to Dr. Guillermo Montoya Blanco (ref: BES-2009-026140)**

A mis padres

A María

## Agradecimientos

Primeramente, me gustaría agradecer a mi director de tesis, Guillemo Montoya, haberme dado la posibilidad de trabajar en su laboratorio y realizar la tesis doctoral en su grupo.

Gracias también a Jasminka Boskovic, no solo por haberme enseñado el mundo de la microscopía electrónica, si no por todos los consejos dados y restaurantes pijos recomendados. Y por supuesto, por haberme aguantado todos los días lado a lado (literalmente). Gracias Jaska

Me gustaría dar las gracias a todos con los que he compartido laboratorio en estos 4 años y pico de tesis. A Inés por todos sus consejos, científicos o no, que me han ayudado mucho en este tiempo. A Pilar por toda su ayuda y disposición. A Rafa, por su humor y por esos kilómetros compartidos corriendo. A Lissy por ser capaz de mantener el orden en un labo lleno de desastres. A Darío, por hacerme reir, y ser capaz de enfadar a Jaska, Ciao Darío!!. A Nehar, por sus consejos y por esos gin-tonics que te han asegurado un sitio en el infierno. A Stefano, por alegrar siempre el labo. A Juán, por ser como es y por enseñarme a manejar los FPLC…A Jesús, el rebelde del laboratorio y con el que se puede discutir de que maratón es el mas rápido. A Igor por suministrarme proteína en los últimos momentos criticos. A Ana, por haber empezado esto juntos y compartir buenos momentos.

Gracias a Juan Méndez y a la gente de su laboratorio. Muy especialmente a Sergio, por su disponibilidad y por enseñarme como funciona eso de las celulitas.

Por supuesto, al resto de la gente del Programa de Biología Estructural. A Ara, por sus consejos y los momentos compartidos, y por esos largos de cuerda que nos quedan por escalar. A Guille (y sus lechones) por hacerme reir y por tu comprensión. A Chevi, y sus métodos científicos “Old School”. A Johanne, por su humor, y por sus brownies, que aunque siguen sin ser perfectos, poco a poco lo

va consiguiendo. Sigue intentándolo ☺. A Marija, que es capaz de resolver estructuras de NMR como pasatiempo. A Alba, por su alegría y por hablarnos de pajaritos y cosas de biólogo (muchas gracias!!!). Paco, mucha suerte en la que te has metido...Maria, que acaba de llegar y espero que consiga quedarse por los madriles.

Lore, vielen Danke für deine Unterstützung. Du hast mich den besten Teil der Wissenschaft geziegt. Irgendwie, du bist die Schludige das ich der Thesis gemacht habe. Danke schön!!!

Muchas gracias a Fer, y su Zona Ruteros. Por esos kilometros compartidos con la bici, y escalando. Y esos bocatas de calamares en Patones!. Al Gominolas-Petazetas team, por esos dias tan buenos en la montaña, y porque con buen humor no existe el mal tiempo. A Luis, por volverme a contagiar el correr por la montaña. Un dia si eso....salimos juntos. A Felix y su locura por el esqui de montaña...

Y por ultimo,

A mis padres. Por enseñarme que el esfuerzo merece la pena. Por enseñarme todo lo que sé y todo lo que me queda por saber. Por su apoyo en todos los proyectos en los que me he metido. En fin, por todo. Muchas gracias.

A Maria, por alegrarme todas las mañanas, tardes y noches. Por escucharme todos los dias, y por estar ahi siempre. Por ofrecerme siempre tu ayuda y tu apoyo. Gracias por hacerme reir tanto, y hacer que los problemas sean mucho mas pequeños. Muchas gracias.

# PRESENTACIÓN

El estudio de las interacciones entre DNA y proteínas nos permite entender los procesos básicos en la célula. La replicación del ADN es un mecanismo esencial para la división y proliferación celular tanto en procariotas como en eucariotas, y está estrictamente regulado a lo largo del ciclo celular. La replicación del ADN implica a muchas proteínas, entre ellas las helicasas. Las helicasas se encargan de desdoblar la doble cadena de ADN usando la hidrólisis del ATP. El complejo heterohexamérico MCM2-7 actúa como la helicasa replicativa en eucariotas y su función es esencial durante las fases de iniciación y elongación de la replicación. La asociación del complejo MCM2-7 con la proteína Cdc45 y el complejo GINS es imprescindible para su función helicasa *in vivo*. Este complejo formado por Cdc45, MCM2-7 y GINS se conoce como complejo CMG y es la verdadera helicasa en eucariotas y forma el núcleo principal del replisoma. Su formación esta altamente regulada por kinasas dependientes del ciclo celular y por las modificaciones post-traduccionales acaecidas.

Hemos co-expresado y purificado el complejo humano MCM2-7 (hMCM2-7) a partir de células de insecto infectadas con un baculovirus recombinante. A través del uso de microscopia electrónica de partículas sueltas y la reconstrucción en 3D, hemos obtenido la estructura a baja resolución del complejo hMCM2-7 unido a ATPγS y del complejo hMCM2-7 unido a overhangDNA en presencia de ATPγS. La presencia del DNA produce unos cambios conformacionales claros en el complejo, haciéndolo a su vez mas estable. La caracterización bioquímica ha demostrado la actividad ATPasa del complejo hmcm2-7 así como su actividad helicasa. Usando el extracto de células HeLa sincronizadas en G1/S, asi como la proteína Cdc45 y el complejo GINS sobre-expresados, hemos conseguido ensamblar el complejo CMG in vitro. Análisis por espectrometría de masas han revelado fosforilaciones necesarias para el ensamblaje del CMG ademas de permitirnos identificar otras proteínas relacionadas con el replisoma.

Por ello, esta tesis ha servido para arrojar luz sobre la estructura del complejo humano MCM2-7, hasta ahora desconocida, y las modificaciones post-traduccionales necesarias para el ensamblaje del complejo CMG.

# ABSTRACT

The study of protein-DNA interactions allows us to understand basic cellular processes. The DNA replication is an essential mechanism for cell division and proliferation in prokaryotes and eukaryotes, and is strictly regulated along the cell cycle. Many proteins are involved in DNA replication, among them the helicases. The helicases unwind the dsDNA in an ATP-dependent manner. The heterohexameric MCM2-7 complex acts as the replicative helicase in eukaryotes, and its function is essential during the initiation and elongation phase of replication. Furthermore, the MCM2-7 requires the interaction with Cdc45 and GINS complex to unwind the DNA *in vivo*. This complex, called CMG (Cdc45,MCM2-7 and GINS) is the truth helicase in eukaryotes and is the core of the replisome. Its formation is highly regulated by cell-cycle dependent kinases and its post-translational modifications are required for the assembly of the CMG complex.

We co-expressed and purified the human MCM2-7 complex (hMCM2-7) from insect cells infected with a recombinant baculovirus. By using single-particle electron microscopy and 3D-reconstruction, we obtained a low-resolution structure of the hMCM2-7 bound to ATPγS and the hMCM2-7 bound to overhangDNA in presence of ATPγS. The presence of DNA produces clear conformational changes in the hMCM2-7 complex, making the complex more stable. The biochemical studies have shown the ATPase activity of the hMCM2-7 complex and its helicase activity. By using synchronized HeLa cell extract at G1/S phase and over-expressed human GINS and Cdc45 we managed to reconstitute the CMG complex *in vitro*. Mass spectrometry analysis revealed essential phosphorilations for the CMG assembly as well as the identification of other proteins associated with the replisome.

Thus, this thesis has served for filling the information gap regarding the structure of the human MCM2-7 complex and the post-translational modifications required for the assembly of the CMG complex.

# Table of contents


1. Introduction……31

1.1 DNA replication……33

1.2 Initiation of replication in eukaryotes……34

1.2.1 Pre-replication complex……34

1.2.2 Pre-initiation complex……36

1.3 Elongation of replication in eukaryotes……38

1.4 Replicative helicases……40

1.5 The Minichromosome Maintenance Complex, MCM……41

1.5.1 Archaea's MCM complexes……42

1.5.1.1 Structural studies of Archaeal MCM complexes……43

1.5.2 BcMCM complex……44

1.5.2.1 Structure of BcMCM……44

1.5.3 Eukaryotic MCM2-7 complexes……45

1.5.3.1 Structural studies of eukaryotic MCM complexes……48

1.6 The CMG complex……49

1.7 The eukaryotic replisome……50

1.7.1 Phosphorilations in the eukaryotic replisome……52

2. Objectives……53

3. Materials and Methods……57

3.1 Protein analysis. General protocols……59

3.1.1 Absorbance ratio 260nm/280nm……59

3.1.2 Electrophoresis in denaturing conditions……59

3.1.3 Western Blotting……60

3.2 MultiBac Expression system……61

3.2.1 Cloning hMCM2-7 in MultiBac Expression System……62

3.2.1.1 hMcm7……63

3.2.1.2 hMcm2 and hMcm4……63

3.2.2 Recombination of vectors for MultiBac Expression System……66

3.2.3 Bacmid DNA preparation……67

3.2.4 Insect cells transfection and hMCM2-7 baculovirus generation……68

3.2.5 Virus titration and small scale expression protocol……69
3.3 Cell culture……70
3.3.1 Sf21 insect cells……70
3.3.2 HeLa cells……71

3.4 Isolation of hMCM2-7 protein complex……71

3.5 Biochemical analysis of the hMCM2-7……72
3.5.4 DNA binding assay……72
3.5.5 ATPase assay……74
3.5.6 Helicase assay……75

3.6 Structural analysis of hMCM2-7……76
3.6.1 Electron microscopy……76
3.6.1.1 Sample preparation and dataset generation……76
3.6.1.2 Two-dimensional images analysis and 3D reconstruction…77
3.6.2 Cryo-electron microscopy……78
3.6.3 Crystallography……78

3.7 CMG reconstitution……79
3.7.1 HeLa cell synchronization……79
3.7.1.1 G1/S arrest……79
3.7.2.2 Mitotic arrest……79
3.7.2.3 DNA staining with propidium iodide……79
3.7.2.4 Preparation of G1/S and mitotic cell extract……80
3.7.2 CMG assembly……80

3.8 Proteomic analysis……81
3.8.1. Sample preparation……81
3.8.2. LC-MS/MS analysis……81
3.8.3. Data analysis……82

4. Results……83
4.1 Cloning of the hMCM2-7 complex……85
4.2 Over-expression and purification of the hMCM2-7 complex.……87
4.3 Biochemical characterization of the hMCM2-7 complex……90
4.3.1 ATPase activity……90
4.3.2 Helicase assay……91
4.3.3 DNA binding assay……93
4.4 Structural characterization of the hMCM2-7 complex……94
4.4.1 2D analysis of hMCM2-7 APO and hMCM2-7 bound to ADP……94
4.4.2 3D reconstruction of the negatively-stained hMCM2-7 bound to ATPγS……95
4.4.3 3D reconstruction of the negatively-stained hMCM2-7 bound to ATPγS and overhangDNA……97
4.4.4 Cryo-electron microscopy. 2D analysis of the hMCM2-7 bound to ATPγS and overhangDNA……99
4.5 The CMG complex……100
4.5.1 hCMG complex assembly……100
4.5.2 Co-expression of hMCM2-7 and hCdc45……102
4.5.3 *In vitro* reconstitution of the human CMG complex……103
4.5.4 Re-purification of human CMG complex……109
4.5.5 Essential phosphorilations for the hCMG reconstitution……110

5. Discussion……113
5.1 3D structure of the human MCM2-7 complex……115
5.1.1 Structure of the hMCM2-7 – ATPγS complex……115
5.1.2 Structure of the hMCM2-7 – ATPγS – overhangDNA complex…116
5.1.3 Comparison of the different hMCM2-7 structures……117
5.1.4 The Mcm2/Mcm5 "gate"……118
5.2 Helicase activity of the hMCM2-7 complex……119
5.3 The CMG complex as the replicative helicase complex……119
5.3.1 Essential phosphorilations for the CMG assembly……121

6. Conclusions……125

7. Conclusiones……129

8. References……133

## Figures Index

Figure 1: DNA replication in eukaryotes……38
Figure 2: Structural studies of the N-terminal of domainof archaeal complex……44
Figure 3: 3D structure of the BcMCM complex……45
Figure 4: Organization of the MCM structural motifs……46
Figure 5: The two conformational states of the *D. melanogaster* MCM2-7 complex……49
Figure 6: *Drosophila* CMG structures……50
Figure 7: Scheme of the multigene transfer vectors and its assembly by Cre recombinase…62
Figure 8: Representation of purification of the hMCM2-7+DNA……74
Figure 9: The pyruvate kinase/lactate dehydrogenase coupled assay used for the measurement of the ATPase activity of hMCM2-7……75
Figure 10: Shared motifs among the human Mcm subunits……85
Figure 11: MultiBac expression system and Bacmid generation……87
Figure 12: Expression and purification of the hMCM2-7 complex……88
Figure 13: Analysis of the hMCM2-7 complex……89
Figure 14: Analysis of hMCM2-7 complex using different gels in denaturing conditions……90
Figure 15: Schematic representation of the couple reaction of lactate generation by PEP and ATP hydrolysis……91
Figure 16: Quantification of ATPase assay……91
Figure 17: Helicase assay of the hMCM2-7……92
Figure 18: Purification of the hMCM2-7 complex together with overhangDNA……94
Figure 19: Reference-free 2D classification of the hMCM2-7 in apo state and 10mM ADP-bound state……95
Figure 20: 3D reconstruction of the human MCM2-7complex with ATPγS……96
Figure 21: 3D reconstruction of the human complex MCM2-7 with overhangDNA……98
Figure 22: Cryo-electron microscopy of the hMCM2-7 complex with overhangDNA……99
Figure 23: Assembly of the CMG complex……101
Figure 24: Assembly of the CMG complex and purification……102
Figure 25: Co-expression of hMCM2-7 complex and human Cdc45 protein……103
Figure 26: Schema of the CMG reconstitution……104
Figure 27: Proteins detected at the CMG reconstitution assay……106
Figure 28: Representation of the levels of PSMs detected for the CMG proteins……107
Figure 29: DNA-damage related proteins……108
Figure 30: Western blot analysis of purified CMG components……110

Figure 31: Phosphorylation in hMCM2-7……..111
Figure 32: Comparison of the hMCM2-7 in presence/absence of DNA……..…117
Figure 33: Representation of the conformational changes of the hMCM2-7 upon ATPγS and DNA binding……..119
Figure 34: Alignment of eukaryotic Mcm2 and Mcm3 and SsoMcm and schema of the hMcm2 and hMcm3……122

# Tables Index

Table 1: Specific antibodies used for detecting the MCM subunits, Cdc45 and the GINS subunit Psf3……61

Table 2: Resume of the restriction sites used for each Mcm subunit……..63

Table 3: Primers used for amplification and cloning of human mcm 2, 4 and 7……..65

Table 4: PCR conditions for the hMcm7, hMcm4 and hMcm2 genes……..65

Table 5: Primers corresponding to the Multiple Cloning Sites 1 and 2 valid for all MultiBac Expression System vectors……65

Table 6: Specific primers designed to verify whole mcm subunit genes……66

Table 7: Resume of clones used for expression of the human complex MCM2-7……..67

Table 8: Oligonucleotides used for DNA binding assays……73

Table 9: Sequences of the oligonucleotides used for the helicase assay……..76

Table 10: Resume of the components used for CMG reconstitution assay……..105

# GLOSSARY

AAA+ ATPase: ATPases Associated with various celular Activities

ADP: adenosine diphosphate

ATP: adenosine triphosphate

ATPγS: adenosine 5′-gamma thiotriphosphate. One of the gamma-phosphate oxygens is replaced by a sulfur atom

*B. cereus*: *Bacillus cereus*

BcMCM: *Bacillus cereus MCM*

bp: base pairs

BSA: Bovine seroalbumine

CDK: cyclin dependent kinases

CMG: Cdc45/Mcm2-7/GINS

C-terminal: carboxyl-terminal extreme

Cryo-EM: Cryo-electron microscopy

*D. melanogaster*: *Drosophila melanogaster*

DDK: Dbf4-dependent kinases

*Dm*MCM2-7: *Drosophila melanogaster* MCM2-7 heterohexamer

DNA: deoxyribonucleic acid

dsDNA: double strand DNA

*E. coli*: *Escherichia coli*

EDTA: ethylendiamintetraacétic acid

EM: Electron microscopy

EMAN: Electron Micrograph ANalysis

FSC: Fourier Shell Correlation

HA: Human influenza hemagglutinin

hCdc45: human Cdc45 protein

HeLa cells: cervical cancer cell line from Henrietta Lacks

hGINS: human GINS complex

His: histidine

hMCM2-7: human MCM 2-7 heterohexamer

IPTG: Isopropil-β-D-tiogalactopyranoside

kDa: kiloDalton

LDH Lactate deshydrogenase

NADH: nicotinamide adenine dinucleotide

Ni-NTA: Nickel-nitrilotriacetic acid

nt: nucleotide

N-terminal: amino-terminal extreme

MCM: Mini Chromosome Maintenance

MkaMCM: *Methanopyrus kandleri MCM*

*M. thermoautotrophicus: Methanobacterium thermoautotrophicus*

MthMCM: *Methanobacterium thermoautotrophicus MCM*

MW: molecular weight

ORC: Origin Recognition Complex

PCNA: Proliferating Cell Nuclear Antigen

PCR: polymerase chain reaction

PEP: phosphoenol-pyruvate

PK: pyruvate kinase

Pol: polymerase

Poli (T): poli-Thymidine

pre-IC: pre-initiation complex

pre-LC: pre-Loading Complex

pre-RC: pre-replicative complex

PSM: peptide spectrum match

RFC: Replication Factor Complex

RNA: Ribonucleic acid

RPA: Replication Protein A

RPC: Replisome Progression Complex

RPM: revolutions per minute

*S. cerevisiae*: *Saccharomyces cerevisiae*

SDS: Sodium Dodecyl sulfate

SDS-PAGE: electrophoresis in denaturing conditions

Sf21: ovarian cells isolated from Spodoptera frugiperda

ssADN: single strand DNA

*S. solfataricus: Sulfolobus solfataricus*

SsoMCM: *Sulfolobus solfataricus MCM*

Strep: streptavidin

SV 40: Simian vacuolating virus 40

TBE: Tris-Borate-EDTA

TEV: Tobacco etch virus

Tm: Mean temperature

Tris: tris (hydroxymethyl) aminomethane

um: micro meter

XMIPP:" X-windows based Microscopy Image Processin Package"

Å: Angstrom

2D: Two dimensions

3D: Three dimensions

# 1. INTRODUCTION

## 1.1. DNA replication

The cell cycle is a sophisticated space-time regulated process leading to genome duplication and cell division. The cell cycle is divided in four main phases: G1 (Gap1), where the cell increases in size and prepares its DNA for replication; S (synthesis), the genetic material duplication stage, where two identical copies from the original DNA molecule are produced; G2 (Gap2), is the phase in between the DNA synthesis and the cell division, where the cell continues growing and finally M phase (Mitosis), the cell division phase where the cell growth stops and its divide in two daughter cells genetically identical.

DNA replication is an essential process that occurs in all organisms. The genome integrity depends on the correct duplication of the genetic material stored in the cells. During genome duplication double stranded DNA is unwound and each strand serves as a template for the copy of its complementary strand. The process ends with two double stranded DNA molecules each containing, an inherited and a newly synthesized strand (Watson & Crick, 1953; Meselson and Stahl, 1958). The genome duplication is highly regulated during the G1 and S phase through different checkpoints to ensure only one replication of the genetic material per cycle (Sclafani & Holzen, 2007). Entry into S-phase is defined by the initiation of DNA replication. In *Saccharomyces cerevisiae* (*S. cerevisiae*) this occurs when the Sic1 inhibitor is degraded following its polyubiquitination by the ubiquitin-protein ligase Stem Cell Factor (SCF). Once the Sic1 is degraded, the S-phase Cyclin Dependent Kinase (CDK) is activated and induces DNA replication by phosphorylating several proteins in the pre-replication complex. The fact that the control of passing these critical points in the cell cycle is mediated by protein degradation ensures that cells proceed irreversibly in one direction through the cycle.

The DNA replication is divided in two different phases: initiation and elongation. Different proteins take part in the initiation phase, some of them recognize sequences or structural motifs presents at the origins of replication while others unwind the DNA and recruit the rest of cofactors needed for the DNA synthesis (Kelman & Hurwitz, 2003; Machida et al, 2005; Mendez & Stillman, 2003). The elongation phase involves the large-scale DNA synthesis by DNA polymerases.

Interestingly, although the DNA replication machinery evolved differently when prokaryotes, eukaryotes and archaea separated million years ago, the fundamental

properties and mechanisms of the DNA replication process are essentially identical among all species (Woese & Fox, 1977). The replication in prokaryotes is simpler. However in eukaryotes, the mechanism is more complex due to the higher number of proteins implicated and the strict regulation of the process.

## 1.2. Initiation of replication in eukaryotes

In eukaryotic cells, DNA replication initiates from multiple replication origins distributed along multiple chromosomes allowing the cells to replicate large genomes in a relatively short periods of time. On the origins of replication, a multi-protein machinery is assembled leading to the generation of two replication forks moving in opposite directions from the origin. In prokaryotes, viruses, archaea, and in the budding yeast *S.cerevisiae,* the origins of replication are determined by a specific sequences. However, in higher eukaryotes, the molecular events to select these regions remain unknown (Yardimci et al., 2010). The basic mechanism of replication initiation is the same for all eukaryotes and occurs in a serial of steps resulting in the formation of two replication forks that will copy the DNA from the origin of replication in opposite directions (Remus & Diffley, 2009; Sclafani & Holzen, 2007).

### 1.2.1. Pre-replication complex

The replication process starts in late G1 when the origin recognition complex (ORC) recognizes and binds onto the origin of replication in an ATP-dependent process, thus labeling the origin of replication. Then Cdc6 associates with ORC and Cdt1 that loads two inactive MCM2–7 helicases onto the dsDNA in ATP depended manner, leading to the formation of a 'pre-replication complex' (pre-RC) also known as 'licensing' (**Figure 1 A**). The hydrolysis of ATP will produce the ejection of Cdt1 from the pre-RC and its conversion to ORC/Cdc6/MCM2-7 complex (OCM complex). Within pre-RCs, the two MCM2–7 molecules encircle dsDNA as an inactive dimer. In S phase, MCM2–7 complexes are "turned on" by S-phase protein kinases and a large number of accessory factors, which together reconfigure MCM2–7 from a dsDNA-binding mode to ssDNA-binding mode. The separation of replication initiation into two temporally distinct steps, helicase loading and activation, allows these events to be differentially regulated in the way that prevents re-replication. Therefore, in S phase, helicases are activated, but licensing is strictly prohibited, owing to a variety of overlapping mechanisms including Cdt1 proteolysis, inhibition of Cdt1 by Geminin

and inhibition of ORC and Cdc6 by cyclin-dependent kinases (CDKs) (Boos, D. et al 2012, Tanaka, S. et al 2013, Yardimici, H. et al 2014).

**ORC complex:** In budding yeast, ORC binds in an ATP-dependent manner to an AT-rich motif called the ARS consensus sequence (Bell, S.P. & Stillman, B. 1992). In higher eukaryotes, ORC binding to DNA is much less specific, and the mechanism of origin recognition is still not known. ORC comprises six subunits (Orc1–Orc6), five of which (Orc1–Orc5) exhibit homology to AAA+ ATPases while also containing DNA-binding sites. Orc1, Orc4 and Orc5 contain functional ATP-binding sites. Interestingly, the binding of ORC to DNA suppresses its ATPase activity, perhaps to avoid futile hydrolysis until pre-RC formation is more advanced (Klemm et al., 2001). The structure of *Drosophila melanogaster* (*D. melanogaster*) ORC complex was elucidated by single particle cryo-electron microscopy (cryo-EM) (Clarey, M.G. 2006), also the *S. cerevisiae* ORC structure was solved by negative staining EM (Chen Z. et al., 2008).

**Cdc6**: Described in yeast by Hartwell in 1973. In G1, Cdc6 binds to the ORC–DNA complex and is required for loading the MCM2-7 complex onto the DNA. Its ATPase activity is crucial for its function, thus mutants lacking Cdc6 ATPase activity are not able to enter into the S phase (Liu J. et al. 2000). Cryo-EM structure of the ORC-Cdc6 complex indicates that all subunits that have been implicated in ATP binding or hydrolysis (Orc1, Orc4, Orc5 and Cdc6) reside on one side. The DNA-binding domains of Orc1–Orc5 are positioned toward the center of the structure, suggesting that DNA is also nestled within the center of the complex (Dueber E.L. et al 2007; Sun, J. et al 2013).

**Cdt1:** Cdt1 and MCM2-7 form a stable complex that docks onto the ORC–Cdc6–DNA complex as a unit forming the OCCM complex. Cdt1 interacts with the C-terminal part of Cdc6 to promote the association of the MCM2-7 complex with the chromatin. On the other hand, cryo-EM structure of the OCCM complex shows that Cdt1 interacts primarily with Mcm2 but also with Mcm5 and Mcm6. Cdt1 dissociates rapidly from the OCCM complex in an ATP-hydrolysis dependent manner promoted by Orc1 and Cdc6, yielding the ORC/Cdc6/MCM2-7 complex (OCM complex) (Remus, D. et al 2009; Gambus, A. 2011).

**Mcm2-7 complex:** The Mini-Chromosome Maintenance proteins 2-7 (MCM2-7) were first discovered in *S. cerevisiae* (Maine et al. in 1984) by doing genetic screenings for genes involved in the plasmid stability. Later, MCM gen was identified in archaea (Kearsey and

Labib, 1984) and in the prokaryotic *Bacillus cereus* in which the MCM gene is encoded as an integrated prophage in the bacterial genome. (McGeoch A.T. and Bell S.D., 2005). However, at the moment there is not further evidence of MCM genes in other prokaryotes. It has been demonstrated *in vivo* (Bell and Dutta, 2002) and *in vitro* (Bochman and Schwacha, 2007) that MCM2-7 complex is the replicative helicase in eukaryotes. After the loading of the complex onto the dsDNA through the Cdt1 and Cdc6-ORC interaction, a second MCM2-7 complex will be loaded forming a dodecamer through the interaction within amino-terminal parts (N-terminal).

### 1.2.2. Pre-Initiation Complex

When the cell enters into S-phase, the Cyclin-dependent Kinase (CDK) and Dbf4-dependent Kinase (DDK) activate the pre-RC. CDK and DDK phosphorylate the MCM2-7 (Lei et al., 2001), remove Cdc6, Cdt1 and promote the interaction of other factors as Mcm10, Cdc45, TopBP1, Sld2, Sld3 and the GINS complex (Remus et al., 2009). These proteins form now the pre- Initiation Complex (pre-IC) (Takeda et al., 2005). First, CDK will remove Cdc6 and Cdt1 allowing MCM10 to link ORC with MCM2-7. CDK will phosphorylate Sld2 and Sld3, Sld3 protein interacts with cofactor Cdc45 and assemble with the MCM2-7 complex; phosphorylated Sld2 together with TopBP1, DNA polymerase ε and the tGINS complex forms pre-Loading Complex (pre-LC) (Johansson et al., 2010; MacNeil, 2010) will load onto the replication site, acting TopBP1 as linker to join Cdc45-Sld3 with the pre-LC, and constituting the pre-IC (**Figure 1 B**). The formation of pre-IC will activate the MCM2-7 due its interaction with GINS and Cdc45. This complex composed by Cdc45, MCM2-7 and GINS assemble the CMG complex and acts as the core of a higher macromolecular complex called Replisome Progression Complex (RPC).

**MCM10:** Mcm10 appears to stimulate activation of the MCM helicase after recruitment and assembly of the CMG complex (Kanke et al., 2012). The Mcm10 protein can oligomerize and its structure forming a ring-shaped hexamer with a large central channel was delucidated by electron microscopy (Okorokov et al., 2007). The inner domains containing DNA binding sites have been crystallized (Warren et al., 2009) but the whole structure at atomic resolution is still not known.

**TopBP1:** The Topoisomease IIβ Binding Protein I was discovered first in humans when searching for proteins interacting with the Topoisomerase IIβ. TopBP1 is present in different

cell processes as DNA replication, cell cycle checkpoints, DNA repair and transcription (Garcia et al., 2005). TopBP1 contains eight conserved phosphopeptide binding BRCA1 C-terminal (BRCT) domains. BRCT domains are versatile modules that form various domain assemblies and are implicated in numerous functions, including protein-protein interactions, phosphopeptide interactions and DNA binding (Leung C.C. et al., 2013).

**Sld proteins:** Discovered in yeast by looking for lethal mutations in yeast Dpb11 mutant strain (Kamimura et al., 1998). In this experiment, the group reported six mutants (Sld1-6) that were analyzed. Sld4 a homolog of Cdc45, Sld1 codifies Dpb3, a subunit of the DNA polymerase ε, Sld2 and Sld6, also known as Drc1 and Rad53 respectively, codifying proteins related to the cell-cycle checkpoint before entering into the S phase. Sld3 is essential for replication and make a complex with Cdc45 in the origins of replication. Sld5 is part of the GINS complex, also essential for the replication (Kamimura et al, 2001; Takayama et al, 2003).

**Cdc45:** It was described in yeast analyzing the cell cycle of cold-sensitive mutants (Moir et al., 1982). Cdc45 is required for the establishment and progression of the DNA replication fork in eukaryotic cells (Labib K. et al., 2000; Bauerschmidt et al., 2007) as well as increase the *in vitro* helicase activity of the MCM2-7 complex (Gambus et al., 2006). The structure of Cdc45 at low resolution, was solved by Small Angle X-ray Scattering (SAXS), showing a model compatible with the crystallographic structure of bacterial RecJ exonuclease (Krastanova et al., 2011; Szambowska et al., 2013).

**GINS:** This heterotetramer complex is composed of the subunits Psf1-3 and Sld5. This complex is essential for replication initiation and cell cycle progression (Kanemaki et al., 2003). The four subunits are conserved among all eukaryotes but they don't have strong relation between them (Takayama et al., 2003). The level of the hGINS complex keep constant along the cell cycle, but is in S-phase when it presents a higher affinity to the chromatin. It also appears to interact with and stimulate the polymerase activities of DNA polymerase ε and the DNA polymerase α-primase complex (Chang et al., 2007). Negative stain EM structure at low resolution of hGINS complex shows a horseshoe shape where ssDNA may fit in the cavity (Boskovic et al., 2007). The crystal structure of the hGINS full-length protein shows a slightly elongated spindle with central hole (Chang et al., 2007).

Finally, the dissociation of Mcm10, TopBP1, Sld2 and Sld3 from the pre-IC and the assembly of Proliferating Cell Nuclear Antigen complex (PCNA) and the DNA polymerases δ and ε by the Replication Factor Complex (RFC) will give rise to the Replisome Progression Complex (RPC) (**Figure 1 C**).

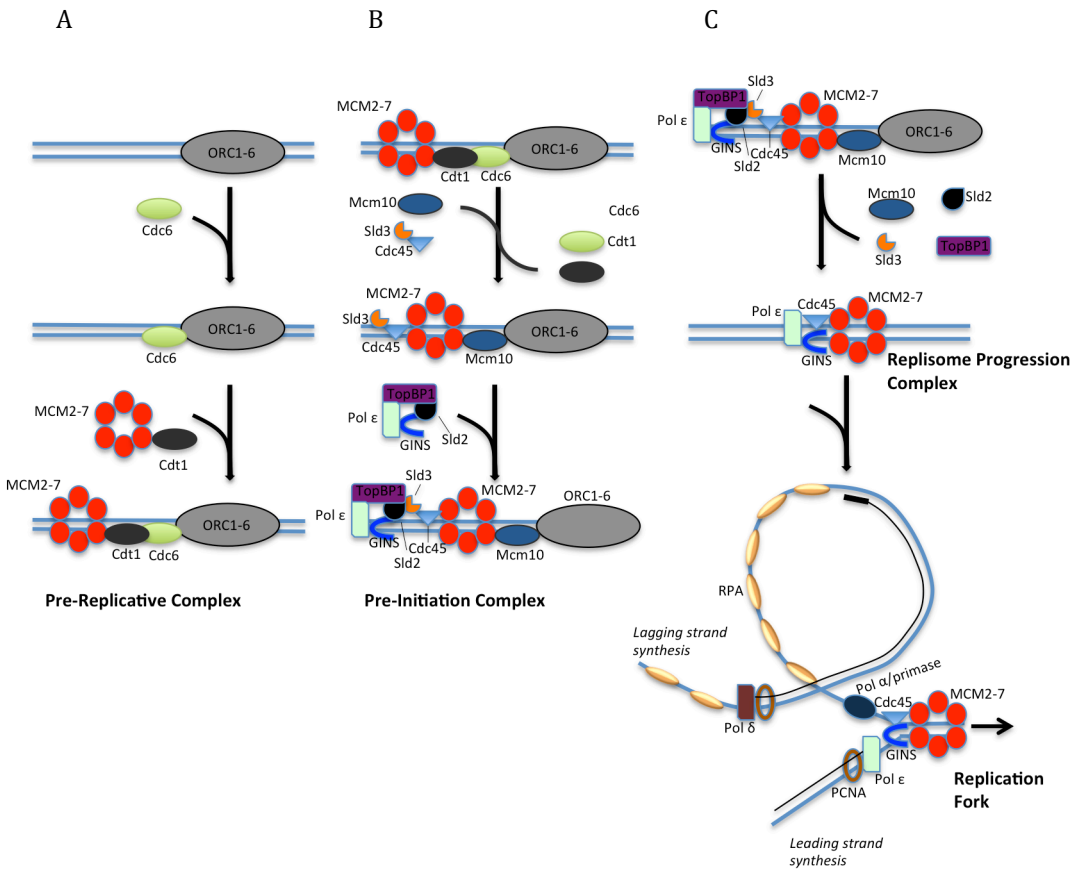



**Figure 1. DNA replication in eukaryotes.** (**A**) Pre-Replicative Complex formation. Once ORC binds to origins of replication, Cdc6 and Cdt1 are responsible of loading the inactive MCM2-7. (**B**) The dissociation of Cdc6 and Cdt1 together with the assembly of other cofactors, among them the pre-LC will form the pre-IC. (**C**) Finally the dissociation of Mcm10, TopBP1, Sld2 and Sld3 give rise to RPC. It starts the dsDNA unwinding following by the recruitment of different proteins that lead to the replisome formation, where the DNA synthesis occurs.

## 1.3. Elongation of replication in eukaryotes

The initiation of replication finishes when the RPC is formed, following the elongation phase where the DNA is synthesized. To start DNA synthesis, the assembly of different proteins at the replication fork is required to build the replisome (Chagin et al., 2010). This replicative complex moves bi-directionally in opposite directions along the DNA together with the replication forks. The DNA double helix is composed of two strands with opposite polarities,

while the DNA polymerase is able to work only in the 5' to 3' direction, thus only one of the two strands will synthesize continuously (the leading strand). The 3'-5' strand, known as lagging stand, will be synthesized discontinuously by small DNA fragments known as Okazaki fragments.

In eukaryotic cells, five different polymerases have been described but only three of them seem to be involved in DNA synthesis, they are the DNA polymerases α, δ and ε. Although these three DNA polymerases have a common catalytic site, each one is specialized in a different function during elongation phase. The DNA polymerase α has primase/polymerase activity, synthetizing small RNA fragments called primers, and extending them by the addition of small DNA fragments (Mossi R. et al., 2000). These primers enable the initiation of DNA synthesis by other DNA polymerases that are not able to synthesize DNA *de novo*. The DNA polymerase δ will synthesize the lagging strand while the DNA polymerase ε will do it on the leading strand (Kunkel & Burgers, 2008; Takeda & Dutta, 2005; Waga & Stillman, 1998).

In the elongation phase, the ORC complex stays at the origin of replication while the CMG complex and the DNA polymerases move with the replication fork. After double helix DNA unwinding, the DNA polymerase α initiates the synthesis of the primers on both strands. At this moment, the five subunits Replication Complex Factor C (RC-F) assembles the PCNA on the dsDNA, which acts as a processivity cofactor for the DNA polymerases δ and ε by surrounding the dsDNA through. These polymerases δ and ε will join to the complex composed by PCNA, the primer and the DNA template. At the same time the DNA polymerase α dissociates from the DNA (Johansson & Macneill, 2010).

Another essential proteins for the DNA elongation process are the Replication Protein A (RPA) polypeptides, which stabilize the single strand DNA generated once the double helix is unbound by the helicase activity of the CMG complex. The topoisomerases are responsible to relieve the torsional stress, generated by the helicase.

The MCM2-7 activity can also be regulated during elongation. The loss of replication fork integrity, an event precipitated by DNA damage, unusual DNA sequence, or insufficient deoxyribonucleotide precursors, can lead to the formation of DNA double-strand breaks and chromosome rearrangements (Lopes et al., 2001; Pasero et al., 2003).

## 1.4. Replicative helicases

The helicases are proteins that use the energy released when hydrolyzing ATP molecules to unwind the helical structure of the double-stranded DNA. The helicases break the hydrogen bonds between the base pairs of the double helix and move along the associated DNA stand in one direction unwinding the complementary strand. Thus, they are also referred as translocases and ATPases-DNA dependent proteins (Tuteja et al., 2004). Because of their essential function, helicases are ubiquitous and evolutionary conserved proteins. They can be found in different states of oligimerzation but most of them are ring-shaped hexameric complexes. The replicative helicases fit DNA into its central cavity, and its ring shape allow them to move along the chromosome (Tuteja et al., 2004). Most of the helicases show DNA binding in an ATP dependent manner. The mechanism of DNA unwinding is one of the main matters of debate in the field. There are two broad types of models that have been proposed: (a) "exclusion" models and (b) "pump" models.

**Exclusion models**. They represent the more traditional view proposing how a hexameric helicase may function. Most hexameric helicases bind ssDNA with higher affinity than dsDNA and can unwind DNA using this method *in vitro*. In this model, a hexameric helicase encircles and translocates along ssDNA towards the fork, unwinding the DNA fork by excluding the opposite strand from the hexamer. The MCM 4, 6, 7 eukaryotic hexamer has been shown to use this mechanism on synthetic fork substrates *in vitro* (Kaplan et al., 2003) and also the archaeal SsoMCM protein (Rothenberg et al., 2007). A variation of the steric model is the ploughshare model (Saikrishnan et al., 2008). This model postulates that the helicase encircles dsDNA and, after local melting of the duplex DNA at the origin, translocates away from the origin, dragging a rigid "wedge" that separates the DNA strands (Takahashi et al., 2005, Singleton et al., 2004).

**Pump models**. In contrast to the exclusion models, recent experiments suggest that hexameric helicases might directly rotate on the DNA. Mechanistic similarities between hexameric helicases and the F1-ATPase, an unrelated hexameric ATPase that couples ATP binding and hydrolysis to the rotation of a component protein within its central channel, have been observed (Shin et al., 2006). Structural work, particularly with SV40 Tag, verifies that the central channel of some hexameric helicases is wide enough to accommodate dsDNA (Kaplan et al., 2003; Kaplan et al., 2004). In addition, six channels of sufficient

diameter to enclose ssDNA lay roughly perpendicular to the main channel and connect the central channel to the exterior of the protein, raising that possibility that ssDNA could be extruded through them (Li et al., 2003; Gai et al., 2004). The rotary-pump model postulates that multiple helicases load at replication origins, translocate away from one another, and in some manner eventually become anchored in place, rotating dsDNA in opposite directions and resulting in the unwinding of the double helix in the intervening region (Laskey et al., 2003; Sakakibara et al., 2009). Experimental evidence consistent with this model includes the finding that pre-RCs may contain up to 50-fold more Mcm2-7 complexes than Orc1-6, a finding that odds with the standard one-helicase-per-fork steric model (Randell et al., 2006; Edwards et al., 2006). Mutation in a region located in the side channel of the SsoMCM inhibits the ATPase and helicase function of the protein. This suggests that the displaced strand could be extruded through the side channels (Brewster A.S. et al., ;MacGeoch A.T. et al., 2005). Similar experiment was carried on with the BcMCM on which the ATPase and helicase levels diminished around the 40% compared with the wild-type BcMCM (Sanchez-Berrondo J. et al., 2011).

The dsDNA pump model is considered as a variation of the pump model, in which two helicases form a head-to-head complex and pump dsDNA toward the origin, where it is extruded as single strands (Bochman et al., 2009).

## 1.5. The Minichromosome Maintenance Complex, MCM

The MCM genes were identified in a genetic screening while looking for the necessary genes for plasmid segregation in S.*cerevisiae* (Maine et al, 1984). Among the MCM gene family, six genes (MCM2-7) are conserved in all eukaryotes, mutations in MCM2, MCM3 and MCM5 in *S.cerevisiae* caused defective plasmid segregation, MCM4 and MCM7 were isolated as cell cycle division mutants, and MCM6 was originally isolated in *Schizosaccharomyces pombe* (*S. pombe*) as chromosome segregation mutant (Maine et al., 1984; Hennessi et al., 1991; Takahashi et al., 1994). Later, the MCM genes were also identified in archaea (Kearsey & Labib, 1998) and in the prokaryotic *Bacillus cereus*, where the MCM gene has a viral origin (McGeoch & Bell, 2005). Archaea contains a single MCM gene that displays higher similarity to the MCM4 than to the rest of the eukaryotic MCM genes (Tye et al., 1999). So far there is no evidence of homologues in other prokaryotes. Curiously, MCM4 and MCM7 were named originally CDC54 and CDC47, whereas MCM6 was known as MIS5. To standardize the

nomenclature the six genes were renamed MCM2 to MCM7 (Chong et al., 1996). The name MCM1 was previously assigned to a transcription factor (Treisman et al., 1992).

### 1.5.1. Archaea's MCM complexes

Several MCM protein homologs have been described in different archaea species. The most characterized are the *Methanobacterium thermoautotrophicus* (MthMCM) and *Sulfolobus solfataricus* (SsoMCM). Several studies suggest that both proteins are not completely identical having certain biochemical and structural differences (Bochman et al., 2009). The archaeal MCM proteins are ATP-dependent helicases, being able to separate the dsDNA with a polarity 3'-5'. The ATP hydrolysis of these proteins is stimulated by the presence of DNA (Jenkinson & Chong, 2006; Sakakibara et al, 2009b). These complexes can bind both ssDNA and dsDNA with different affinities.

Analysis of DNA binding of the SsoMCM truncation mutants, consisting of only the N- or C-terminal domains, suggest that the change from double to single stranded DNA occurs at buried point in the central channel of the enzyme. It seems likely the two single strands would emerge from different cavities leading to the strand extrusion model. It was found that N-terminal-SsoMCM binds ssDNA with an affinity comparable to wild type, but the C-terminal SsoMCM displays a severely affected ssDNA binding (Pucci et al., 2007). It has been reported as well that C-terminal-SsoMCM can bind dsDNA with good affinity but N-terminal SsoMCM cannot (Liu et al., 2008). The N-terminal-SsoMCM structure bears this out, as the central channel seems too narrow to encircle dsDNA. The SsoMCM prefers substrates that contain segments of ssDNA and dsDNA, such as replication fork or bubble structures. This helicase is able to untwist dsDNA as long as it contains a single strand end (Pucci et al, 2004; Rothenberg et al, 2007; Barry et al., 2007). However, the N-terminal double hexamer of MthMCM can bind dsDNA and ssDNA (Fletcher et al., 2003). The MthMCM is able to unwind any DNA structure even those with blunt-ends, but always in the presence of magnesium and ATP (Shin et al., 2003). An EM study indicates that the MthMCM can accommodate both dsDNA and ssDNA in the C-domain and the N-domain, but it may have a preference for dsDNA at the C- domain and for ssDNA at the N-domain (Costa et al., 2006). During DNA unwinding by a single hexamer, if the C-terminal domain binds dsDNA and the N-terminal domain binds ssDNA, then the other ssDNA strand may be extruded between the two domains through one of the six side channels near the ATP binding site. The other strand would continue in the central channel and exit through the N-terminal central channel.

The MCM models from archaea are still used to study features of the eukaryotic MCM2-7 complex due to their simplicity. The only crystal structures solved with atomic resolution from the MCM proteins are the ones coming from archaea (Fletcher et al., 2003; Liu et al., 2008).

### 1.5.1.1. Structural studies of Archaeal MCMs

Structural studies performed on archaeal MCM complexes have been used as models to study the mechanisms of helicase activity. The first structure solved by X-ray diffraction was the N-terminal part of MthMCM (**Figure 2 A**). This fragment forms a double hexamer with a central cavity large enough to fit both ssDNA and dsDNA (Fletcher et al., 2003). The MthMCM complex is able to bind dsDNA around the molecule, suggesting an initial site of interaction between protein and DNA before the loading of the helicase onto the replication fork (Costa et al., 2008). The structure of the near full-length MCM protein from *Methanopyrus kandleri* (MkaMCM) was solved by crystallographic methods. This MkaMCM was isolated and crystallized as a monomer, it has a natural deletion of a critical zinc-binding subdomain in the N-terminal domain, which prevents hexamerization and thus helicase activity (Bae et al., 2009). Also the near full length SsoMCM protein has been crystalized as monomer and solved at 4.9 Å, and the hexameric model was generated by docking of six SsoMCm monomers within the double hexameric volume of MthMCM. (Brewser et al., 2008). The full lenght hexameric model proposed for SsoMCM presents certain clashes in their structure. However, the N-terminal domail hexameric structure contains a central cavity where only ssDNA would fit (**Figure 2 B**) (Liu et al., 2008). Both MkaMCM and SsoMCM have two kinds of cavities; the central one positively charged where DNA locates, and six smaller cavities, at the interfaces of the subunits, around the molecule with a possible role in the unwinding mechanism (Brewster et al, 2008; Costa et al, 2006a).

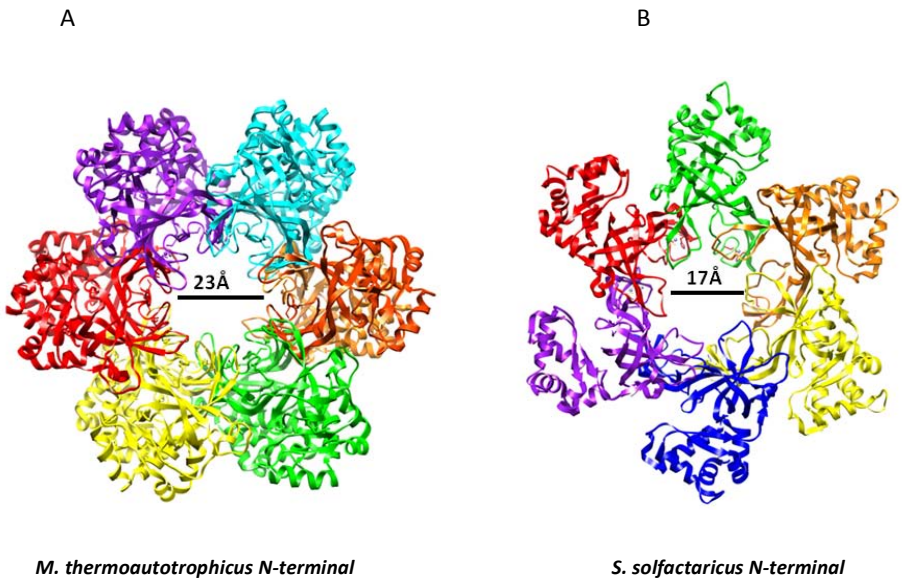



**Figure 2. Structural studies of the N-terminal domain of archaeal MCM complex.** (**A**) Crystallographic structure of the N-terminal domain of *M. thermoautotrophicus* MCM complex (PDB 1LTL) (Fletcher et al., 2003). (**B**) Crystal structure of the S. solfactaricus (PDB 2VL6) (Liu et al., 2008). The differences in the cavity sizes are indicated.

### 1.5.2. BcMCM complex

*Bacillus cereus* contains a MCM gene (BcMCM) that was found as an integrated prophage in the bacterial genome. The N-terminal fragment of the protein contains a region that is homologous to the catalytic subunit of the archaeal–eukaryotic DNA primase (MacGeoch A.T. et al, 2005). The N-terminal domain also shows homology with the primase–polymerase domain of the replication protein ORF904 of plasmid pRN1 from *Sulfolobus islandicus*, which is also fused to a helicase domain (Lipps G. et al., 2004). The BcMCM C-terminal section is homologous to the MCM AAA+ helicases, with typical Walker A and Walker B motifs in the ATP-binding site, having between 20-24% similarity to the human MCM2-7 subunits. The structure and function of BcMCM homohexamer has been characterized (Sanchez-Berrondo J. et al., 2011). The BcMCM shows an ATP-dependent helicase activity performed by its C-terminal domain. In addition to the helicase activity, the primase-polymerase activity arising from the N-terminal domain can initiate the DNA synthesis by using dNTPS as substrate. BcMCM preferentially binds ssDNA than dsDNA. The C-terminal helicase domain (BcMCM501-1028) is sufficient for DNA binding whereas the N-terminal primase domain (BcMCM1-361) do not bind DNA, being the residues 361-400 the responsible for DNA binding in the primase-polymerase domain.

#### 1.5.2.1. Structure of BcMCM complex

The BcMCM is isolated as monomer, however, the addition of nucleotide induces

hexamerization of the protein *in vitro*. The 3D EM structure of BcMCM, at 36 Å resolution, shows a homohexameric complex similar to the archaeal MCM complexes, with a central channel large enough to accommodate dsDNA and six lateral channels enough to allow ssDNA pass through. The complex has been modeled in presence of ATP/ADP and in presence/absence of ssDNA. The nucleotide binding does not alter significantly the structure of the BcMCM (**Figure 3 A**), but in presence of ssDNA introduces large conformational changes. A different conformation has been observed between the BcMCM–ATPγS–ssDNA and the BcMCM–ADP–ssDNA structures in the diameter of the aperture located at the bottom part of the central channel of the complex. This entrance to the central channel is significantly narrower in the BcMCM–ATPγS–ssDNA complex, suggesting that it opens-up upon nucleotide hydrolysis (**Figure 3 B**) (Sanchez-Berrondo et al., 2011). The primase-polymerase domain of the complex is highly flexible, thus it was not visible in the structure (Sanchez-Berrondo J. et al., 2011).

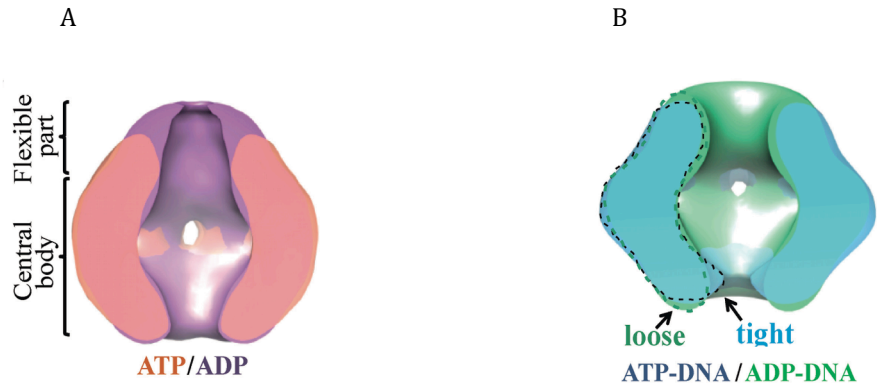



**Figure 3. 3D structure of the BcMCM complex.** (**A**) Superposition of the symmetrized cut-open side views of BcMCM–ADP model and the symmetrized BcMCM– ATPγS. ATP hydrolysis does not introduce large conformational changes in the BcMCM central body region. (**B**) Superposition of symmetrized cut-open side views of BcMCM–ATPS– ssDNA model and the symmetrized BcMCM–ADP–ssDNA. One monomer of a BcMCM– ATPγS –ssDNA is highlighted with a black line while one monomer of the BcMCM–ADP–ssDNA is indicated with a dashed green line (Adapted from Sanchez-Berrondo J. et al., 2011).

### 1.5.3. Eukaryotic MCM2-7 complex

The six proteins that form the heterohexameric complex MCM2-7 contain sequence differences between them, however, in all of them the 250 amino acids AAA+ ATPase domain is conserved. The complex is involved in the initiation and elongation of DNA replication, being each of the six proteins essential for cellular viability (Forsburg et al., 2004; Labib et al., 2000; Tye et al., 1999). *In vivo* experiments have described the MCM2-7

complex as the replicative helicase in eukaryotic cells (Bell and Dutta, 2002) but it was few years later when the helicase activity of the MCM2-7 complex was tested *in vitro,* showing the dependence of the activity in the presence of certain anions, such as acetate or glutamate (Bochman and Scwacha, 2007; Bochman and Schwacha 2008). However, this oligomeric form has historically lacked *in vitro* helicase activity. Instead, a specific dimeric heterotrimer (Mcm4/6/7) was isolated from a variety of systems that showed ATPase and 3'-5' DNA-unwinding activity (Ishimi et al., 1997; Kaplan et al., 2003). Furthermore, the addition of Mcm2 or the Mcm5/3 dimer in the reaction inhibits Mcm4/6/7 helicase activity (Lee and Hurwitz, 2001).

Except for the eukaryotic Mcm2-7 complex, all currently known hexameric helicases are homohexamers. As molecular motors, helicases unwind nucleic acids by coupling the conformational changes caused by nucleotide binding and hydrolysis to the physical manipulation of the nucleic acid. Thus, the study of ATP hydrolysis provides important mechanistic clues about helicase function. As most helicases exhibit ATP-stimulated ssDNA binding, ATP binding apparently causes a conformational change that facilitates DNA binding. Each of the subunits of the eukaryotic MCM2-7 complex contains an ATP-binding site. This ATP catalytic center is found in the interface between two subunits; one of the subunits contains the Walker A, Walker B and Sensor I motifs in a cis-position whereas the second subunit contains the Sensor II and Arginine-Finger motifs in trans-position. The Walker A motif binds the ATP molecule while the Walker B and Sensor I orientate the water molecule for the hydrolysis of the ATP, the Sensor II contacts with the γ-phosphate of the ATP and the Arginine-Finger motif is the responsible for transmitting the information to the next subunit (**Figure 4**) (Bochman and Schwacha 2009).

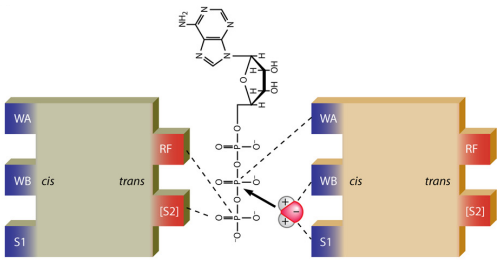



**Figure 4. Organization of MCM structural motifs.** Graphic representation of the ATPase catalytic center located between two MCM subunits. Functional motifs and their functions are represented with discontinued lines (Adapted from Bochman and Schwacha, 2009).

In 2007 Moreau and colaborators. tested the ATPase ctivity of SsoMCM wild type and ATP hydrolysis mutants mixed in different proportions. These data showed a semi-sequential model for the archaeal SsoMCM ATP hydrolysis. Specifically, up to three mutants can be tolerated, as long as the remaining three subunits are sequential. This implies that only half of the hexamer is needed to unwind DNA. If only a few ATP sites are necessary, the rest of the subunits may have no evolutionary pressure in terms of helicase function and can evolve to specialize (Brewster et al., 2010). This point may bear some relevance to the organization in eukaryotic MCMs with six different subunits (MCM2–7): dimers of 7/4, 4/6 and perhaps 3/7 have ATPase activity while the others do not. However, although the other subunits are not required for *in vitro* helicase activity, the corresponding alleles are lethal *in vivo*. Thus, ATP binding and/or hydrolysis at these sites still performs an essential function, possibly to regulate the loading, activation, or unloading of MCM2-7 during the corresponding phases of DNA replication (Bochman and Schwacha, 2009). Mutations in the Walker B or arginine finger motif or the inclusion of just a single Walker A mutant subunit strongly inhibits ATP hydrolysis by the entire Mcm2-7 complex, suggesting that even active sites that demonstrate little or no ATP hydrolysis still act as allosteric activators of ATPase activity (Bochman and Scwacha 2010).

Finally, the MCM 2/5 interface represents an ATP-dependent discontinuity in the MCM2-7 toroidal structure. This interface forms a "gate" that regulates helicase activity and allows the loading onto dsDNA (Bochman and Schwacha, 2008). The MCM2/5 active site functions as an allosteric regulatory site of the MCM2-7 complex. Mcm2-7 complexes bearing mutations that perturb the gate function: Mcm2RA (arginine finger) and Mcm5KA (Walker A box) mutations, lack helicase activity. This suggests the inability to close the gate (Mcm5KA mutation) or a defect in the ability to communicate the state of the gate closure to the rest of the complex (Mcm2RA) blocks helicase activity. In addition, the two active sites flanking Mcm2/5, the Mcm5/3 and Mcm6/2 sites, appear to be involved in coupling ATP hydrolysis and DNA unwinding, as Mcm2-7 complexes that contain Walker A mutations in both sites demonstrate high levels of ATP hydrolysis but little DNA unwinding. These two sites may transmit a positive signal to the rest of the subunits when the gate 2/5 is closed to switch on the helicase activity. The opposite situation might be also true, when the Mcm5/3 and Mcm6/2 transmit a negative signal when the gate is opened, turning off the helicase activity.

The helicase activity of the MCM2-7 was demonstrated *in vitro*, being anion dependent

(Bochman-Schwacha, 2008). *In vivo* ATP levels are high enough to suggest that the Mcm2/5 ATPase active site may normally be filled with ATP. If the Mcm2/5 site functions as the DNA access point in chromosome loading during pre-RC formation, than one possible role of the Cdc6 and Cdt1 would be to change the state of the Mcm2/5 gate from a closed to an opened conformation. Moreover, since MCM2-7 is functionally inert within the pre-RC, the idea that Cdc45 and GINS complex were cofactors for the MCM2-7 was supported by. Experiments with *Drosophila melanogaster* MCM2-7 (DmMCM2-7) and CMG complexes demonstrated that the association of MCM2-7 with Cdc45 and GINS increase ATP hydrolysis and its affinity to bind DNA (Ilves et al., 2009).

### 1.5.3.1. Structural studies in eukaryotic MCM complex

The six proteins that compose the MCM2-7 complex, interact forming a ring-shape complex with a stoichiometry of 1:1:1:1:1:1 and a determined position within the ring (Davey et al., 2003). The crystallographic structure of the eukaryotic MCM proteins is unknown so far. Therefore all the structural information available arises from low- resolution EM analysis. These studies have shown the nearly six-fold symmetry of the ring with a diameter of 145Å and a central cavity of 25 Å-30 Å (Sato et al., 2000; Bochman and Schwacha, 2007).

In budding yeast, the MCM2-7 complex can form double-hexamers in an ATP-dependent manner on dsDNA in presence of ORC complex, Cdc6 and Cdt1. The double-hexamer model solved at 30 Å resolution shows the protein bridge at the inner and outer circumference of the N-terminal rings, are slightly different from the MthMCM, where the interactions between the two MCM molecules are mainly at the inner circumference of the N-terminal rings (Remus et al, 2009).

Structural studies have shown the Mcm2/5 gate in MCM2-7 from *D. melanogaster*, and have supported the idea of the conformational changes in the molecule, from a more opened and flexible conformation in which the Mcm2 and Mcm5 are well separated (Spring-lock form) to another tighter and more rigid conformation with a smaller separation between Mcm2 and Mcm5 (Notched-planar form). These two architectures seem to exist in equilibrium in the apo state, even in presence of ADP-BeF3 (**Figure 5 A-B**) (Costa et al., 2010). The positions of the different subunits have been shown by biochemical methods and by comparing class averages from images obtained by EM of the N-terminal domain of the different MCM subunits labeled with MBP (Costa et al., 2010).

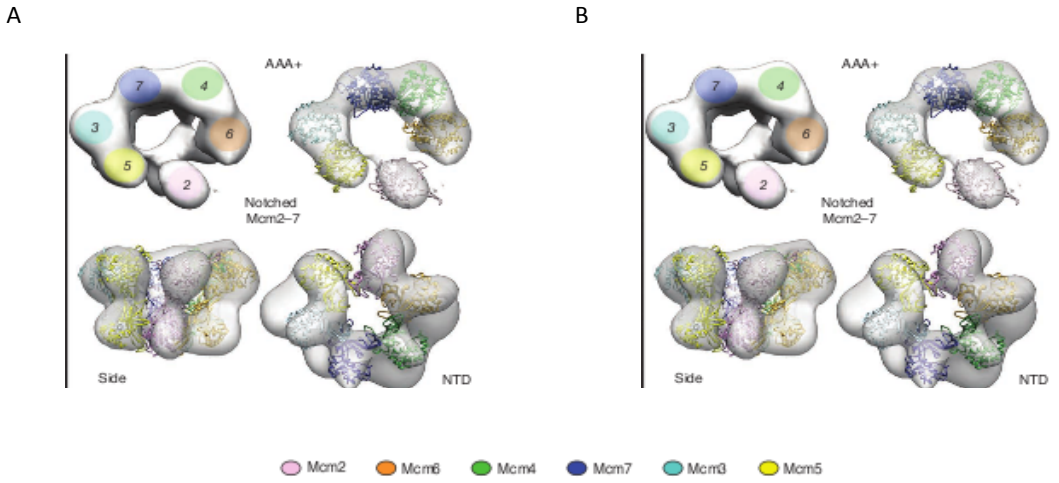



**Figure 5. The two conformational states of the *D. melanogaster* MCM2-7 complex.** (**A, B**) Reconstructions of the notched-ring or lock-washer models of *Dm*MCM2-7 complex. The different MCM subunits are represented in different colors. Homolog SsoMCM crystal structure has been fitted into the volume (Adapted from Costa et al., 2011).

## 1.6. The CMG complex

CMG complex is composed by Cdc45, MCM2-7 and GINS, building the core of the replisome. The MCM2-7 complex by itself has a little helicase activity *in vitro*. Therefore, it was hypothesized the need for cofactors that would activate the helicase. The first evidence of the CMG complex was obtained from *Xenopus laevis* egg extract (Kubota el al., 2003). Similar studies were carried on in other organisms, such as budding yeast or fruit fly confirming the importance of this complex and propose it as the real helicase *in vivo* (Moyer et al., 2006; Aparicio et al., 2006).

It is known that CMG complex is essential for initiation and elongation of replication, as well as being part of the replisome (Gambus et al., 2006). The CMG complex from *D. melanogaster* has been co-expressed *in vitro* by Ilves et al., 2010, and showed how the ATPase activity of the MCM2-7 increased in presence of Cdc45 and GINS by two orders of magnitude, together with the improvement in the helicase activity and the recognition of DNA as substrate. As the MCM2-7, CMG complex cannot bind dsDNA (Remus et al., 2009; Ilves et al., 2010).

There is still little information about the structure of the CMG complex. In Costa et al., 2011 were able to co-express in insect cells and purify the CMG complex from *Drosophila*, and use it to obtain a low-resolution structure (28Å-35Å) that offers important information on the

locations of the CMG components. As previously mentioned, the MCM2-7 may co-exist in two different conformations, planar or spring, however, the MCM2-7 in the apo CMG presents only the planar conformation. Moreover, the gap between mcm2/5 is still present at the apo CMG complex, and is considerably narrower than the gap present in the ADP-$BeF_3$ MCM2-7 (**Figure 6 A**). The CMG in presence of ADP-BeF3 showed structural features similar to the apo CMG complex, although the C-terminal domain of the MCM2-7 subcomplex is considerably more constricted closing the gate and narrowing the diameter of the pore (**Figure 6 B**). By docking the crystal structure of human GINS complex that shares around the 40% sequence identity on average with each of the four *Drosophila* subunits, Costa et al., 2011 could observe the extensive contacts of GINS with the N-terminal domain of Mcm3 and Mcm5 through the Psf2 and Psf3. The unfilled density after docking may correspond to Cdc45 protein. In this model, Cdc45 would share a large interaction surface with the N-terminal α-helical domain of Psf2 and associates with the N-terminal domain of Mcm2 (Im J.S. et al., 2009; Costa et al., 2011). This assignment demonstrates that the GINS-Cdc45 subcomplex forms a handle that bridges the Mcm2/5 gap (Ilves et al., 2010; Costa et al., 2011).

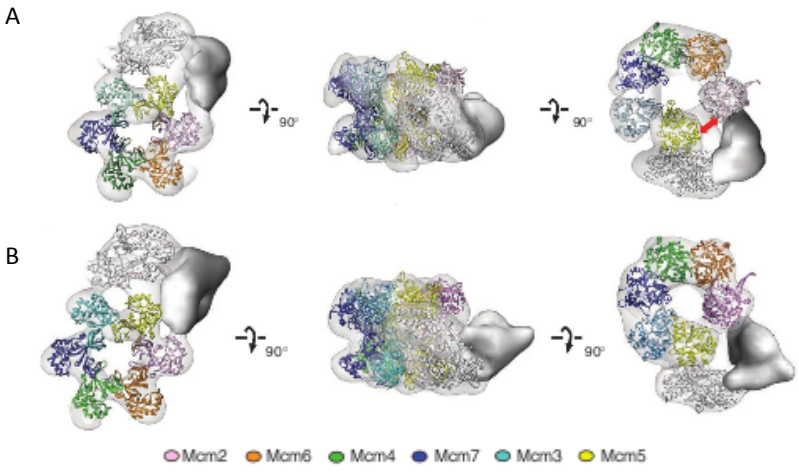



**Figure 6. Drosophila CMG structures.** (**A**) apo CMG structure. From left to right views of the structure: N-terminal domain, side view and C-terminal domain where the gap between mcm2/5 is labeled with a double arrow. (**B**) ADP-BeF3 bound to CMG. Homolog SsoMcm was used for docking. The different Mcm subunits are represented in different colors whereas GINS is colored in white and Cdc45 as a solid gray (Costa et al., 2010).

## 1.7. The eukaryotic replisome

The proteins required for moving the replication fork act together as a machine referred to as the replisome. The basic enzymatic activities of cellular replisomes are common to all domains of life and include the DNA polymerases, proofreading 3'-5' exonucleases, a

hexameric helicase, primase, and a clamp loader that assembles ring-shaped sliding clamps onto primed sites to tether polymerases to DNA for high processivity (Kurth et al., 2013). Although the conservation of these replisome components is substantial, this is where the similarity between eukaryotic and prokaryotic replisomes stops. The eukaryotic replisome is composed at the moment by 48 polypeptides although many additional factors have not been yet described. As well as the well described replisome components CMG, Pol ε, Pol δ, the primase Pol α, PCNA or the Replication Factor C (RFC) clamp loader, recent studies have identified proteins interacting with the CMG complex forming what is known as Replication Progression Complex (RPC) putting in additional components to the replisome, as Mcm10, Ctf4 (chromosome transmission fidelity 4), Mrc1 (mediator of the replication checkpoint 1), and possibly other proteins (Im J.S. et al., 2009; Gambus et al.,2009). Many of these proteins lack homologs in bacterial systems (e.g., Cdc45, GINS, and all accessory subunits of Pols α, δ and ε).

The complexity of the eukaryotic replisome relies also in its high regulation by several post-translational modifications, even DNA damage introduces additional components to the replisome (Kurth et al., 2013). Difficulties in the progression of the replication fork lead to the accumulation of ssDNA, which produces deletions, duplications, or complex genome rearrangements. This accumulation of ssDNA at replication forks is known as replicative stress and has been associated to cancer and aging (Lopez-Contreras et al., 2010). In addition to those factors directly involved in DNA replication, proteins involved in DNA repair, cell-cycle checkpoints, or chromatin remodeling are also enriched in the proximity of the replisomes (Maga et al., 2001). One of the key targets for regulation is PCNA, which can serve as a binding platform for various enzymes involved in DNA repair, chromatin assembly, and cell cycle control (Fox J.T. et al., 2011). Also the Mismatch Repair system (MMR) is related with DNA replication (Flores-Rozas et al., 2000; Jiricny, 2006; Lopez-Contreras et al., 2013). MMR is composed by the MutL complex, the Msh2-Msh6 proteins and by the much less abundant MutS complex, the Msh2-Msh3 proteins (Hombauer et al., 2011). The function of this complex is to eliminate mispaired bases resulting from replication errors by the recruitment of the Mlh1-Pms1 complex. An interaction between PCNA and the MMR proteins MSH6 and MSH3 has been probed (Clark et al., 2000), as well as the enrichment of Msh2 on precipitated nascent DNA. All these data demonstrates that these proteins are either directly linked to the replisome or are abundant in the vicinity of the replisome (Lopez-Contreras et al., 2013).

### 1.7.1. Phosphorilations in the eukaryotic replisome

The conserved kinase Cdc7 acts in association with an essential regulatory subunit called Dbf4. Cdc7 is also referred as DDK (Dbf4-Dependent Kinase). It was found in budding yeast that Dbf4–Cdc7 interacts with Mcm2, phosphorylates Mcm2 *in vitro*, and is required for Mcm2 phosphorylation *in vivo*. Furthermore, it has been proved that Mcm3, Mcm4, and Mcm6 are also phosphorylated by Cdc7–Dbf4 *in vitro*, and a later studies showed that similar occurs for Mcm7 (Lei et al., 1997; Weinreich and Stillman 1999). In contrast, Mcm5 has not been found to be a substrate of Cdc7–Dbf4, in budding yeast or any other species. Interestingly, it looks like the major phosphorylation sites for Cdc7-Dbf4 and CDK reside in the N-terminal of Mcm2, Mcm4 and Mcm6 (Sheu and Stillman 2006) and these phosphorylations in MCM2-7 complex are enhanced when Cdc45 is present (Sheu and Stillman 2010) suggesting the improvement of the stability of the Cdc45-MCM2-7 complex. Whereas DDK looks the main kinase for the MCM2-7 complex (Labib et al., 2010), CDK seems to be the responsible to phosphorylate Sld2 and Sld3, cofactors needed for the recruitment of GINS and Cdc45 respectively (Zegerman et al., 2007).

In vertebrate cells there are evidences that N-terminal part of Mcm2 and Mcm4 are phosphorylated by Cdc7-Dbf4. Cdc7 has a strong preference for Ser or Thr in an acidic context, either produced by adjacent amino acids that are acidic, or elsewhere due to an adjacent CDK site that appears to prime phosphorylation by Cdc7, meaning that the phosphorylation produced by the CDK may help the phosphorylation of the neighbor residue by DDK (Olsen et al., 2010). Using *Xenopus* eggs extracts was demonstrated that Cdc7 phosphorylates the MCM2–7 complex preferentially on chromatin during S phase (Jares and Blow 2000). The consequences of the phosphorylation of the MCM2–7 helicase by Cdc7 in human cells also seem to be similar to the situation in yeast, Cdc7-Dbf4 and CDK are required for the formation of the CMG helicase complex in human cells (Im et al. 2009). Clearly there is still much to learn regarding how the MCM2–7 helicase and CMG formation are regulated by DDK and CDK phosphorylation in human cells.

# 2. Objectives

The aim of this thesis is to analyze the structure and function of the human replicative helicase MCM2-7 complex and to shed light on the assembly of the replisome core, CMG complex. In particular, the following objectives have been pursued:

1- Overexpression and purification of the human MCM2-7 complex

2- Functional characterization of the complex through biochemical and biophysical studies.

3- Three-dimensional reconstruction of the human MCM2-7 complex in presence and absence of DNA.

4- Reconstitution of the CMG complex.

# 3. Materials and Methods

## 3.1. Protein analysis. General protocols

These protocols were applied along the thesis for protein characterization

### 3.1.1. Absorbance ratio 260/280nm

To test the possiblity of the presence of DNA in our sample, the absorbance ratio at 260 nm/280 nm was evaluated in a spectrophotometer NanoDrop (ND-1000) by loading 2 ul of the protein solution at room temperature. The absorbance of the sample was measured in a wide spectrum of wave lengths from 220 nm to 350 nm.

### 3.1.2. Electrophoresis in denaturing conditions

Electrophoresis in denaturing conditions was done in different Mini-Protean Precast Gels at 4%-15%, 4%-12% polyacrilamide gradient concentration and at continuous 12% polyacrilamyde concentration or at AnyKD gel (Bio-Rad) or in NuPage 7% Tris-Acetate Gel (Invitrogen). Before loading onto a gel the protein solution was mixed with 6x Loading Dye (125 mM Bis-Tris pH:6.8, 20% v/v glycerol, 4% w/v Sodium Dodecyl Sulfate, 10% w/ β-mercaptoethanol, 0.4 mg/ml bromophenol blue) and boiled for 3 minutes at 95 °C . For protein visualization, the gels were stained with Coomassie Blue (Simply Blue Safe Stain, Invitrogen) to check the presence of the complex proteins. The molecular weight standards corresponds to the SeeBlue-Plus2 (Invitrogen) or Precision Plus Protein Standards (Bio-Rad). The gels were run in 1x Electrophoresis buffer (25 mM Tris-Cl, 250 mM glycine, 0.1% SDS) at 180 V. In some cases the gels were run in MOPS buffer (40 mM MOPS, 10 mM Sodium Acetate and 1 mM EDTA) or in TAE buffer (40 mM Tris-Acetate and 1 mM EDTA).

The complex was also analyzed in home made 20 cm SDS-PAGE. For that the stacking gel was done at 8% acrylamide with Tris-HCL and pH:6.8. The concentration of acrylamide in the running gel was 8% in 370 mM Tris HCL at pH: 8.8. The final concentration of SDS, in the two parts of the gel, was 0.1% with 1 mm gel thickness. The polymerization reaction catalyst for both gels were 0.01% TEMED (N,N,N,N'-trimetilnediamine) and 0.1% PBS (amonium persulfate). These gels were run in electrophoresis buffer at 100 V and stained with Simply Blue Safe Stain (Invitrogen).

In some cases, the protein was visualized by silver staining immersing the gel in Fixing solution (200 ml EtOH, 50 ml acetic acid and water to 500 ml) for 30 min, then in Sensitizing solution (150 ml EtOH, 2.5 ml glutaraldehyde, 1 g sodium thiosulphate, 34 g sodium acetate and water to 500 ml) for 30 min followed by three washes in 100 ml water for 5 minutes each. Later, the gel was inmeresed in Silver solution (0.5 g silver nitrate, 200 ul formaldehyde and water to 500 ml) for 20 min, washed quickly with water and transfered into Developing solution (12.5 g sodium carbonate, 100 ul formaldehyde and water to 500 ml) until protein bands appeared. The reaction was stopped by immersing the gel in Stop solution (7.3 g EDTA and water to 500 ml).

### 3.1.3 Western Blotting

The proteins of the complex hMCM2-7 were analyzed by Western Blot using specific antibodies for each hMcm subunits kindly provided by the DNA-Replication Group at the CNIO (Madrid, Spain). Aliquots of specific antibodies for HA-tag and FLAG-tag and immunofluorescence secondary antibodies were provided by the Cell Division and Cancer Group at the CNIO (Madrid. Spain) (**Table 1**).

The proteins of the complex were separated by SDS-PAGE and then transferred onto nitrocellulose membrane previously soaked in Transfer buffer (3 g Tris, 14.4 g glycine, 200 ml MetOH and water to 1 L,pH:8.6). The protein transfer was done in a wet system by covering completely the membrane and gel in a transfer buffer and transfering at 100 V for one hour at 4 °C . After that the membrane was soaked into 10 ml Blocking solution (1 g powder non-fat milk, 50 ul Tween and PBS up to 100 ml) for one hour at room temperature. Then, 10 ml of Blocking solution was added with the specific antibody in dilution 1:1000 and incubated for 3 hours at room temperature.

| *Antibody* | *Host specie* | *Dilution* |
|---|---|---|
| Anti-Mcm2 | Rabbit | 1:1000 |
| Anti-Mcm3 | Rabbit | 1:1000 |
| Anti-Mcm4 | Rabbit | 1:1000 |
| Anti-Mcm5 | Rabbit | 1:1000 |
| Anti-Mcm6 | Rabbit | 1:1000 |
| Anti-Mcm7 | Rabbit | 1:1000 |
| Anti-HA | Mouse | 1:1000 |
| Anti-FLAG | Mouse | 1:8000 |
| Anti-His | Rabbit | 1:5000 |
| Anti-Strep | Mouse | 1:5000 |
| Anti-Cdc45 | Mouse | 1:1000 |
| Anti-Psf3 | Mouse | 1:1000 |

**Table 1. Specific antibodies used for detecting the MCM subunits, Cdc45 and the GINS subunit Psf3.**

After washing the membrane with 1x PBS+0.05% Tween20, 10 ml of blocking solution with the secondary antibody 1:1000 goat anti-rabbit HRP conjugate for anti-Mcm proteins or 1:2000 anti-mouse alkaline phosphatase conjugate for HA-tag and FLAG-tag antibodies were added. The membrane was incubated for one hour at room temperature and washed again three times with 1x PBS+0.05% Tween20.

For developing the protein bands, the membrane was incubated following the protocol of ECL Prime Western Blotting Detection Reagent (GE Healthcare), the blot is then exposed on X-ray film and developed.

## 3.2 Multibac Expression System

The Multibac Expression System is a modular baculovirus-based system specifically designed for the expression of large eukaryotic multiprotein complexes developed by the Ritchmond lab at ETH in Zürich (Switzerland) (Fitzgerald Daniel J. et al., 2006).

The MultiBac Expression System is composed of different transfer vectors which facilitate the generation of multigene cassettes for protein co-expression, thus providing a flexible platform for generation of protein co-expression vectors (Fitzgerald Daniel J. et al., 2006). This system has two types of vectors; acceptor vector (pFL) and donor vectors (pSpL and pUCDM). Derivatives of one acceptor plasmid can be fused to one or two donor plasmids creating multigene plasmid dimers or trimers, respectively. The assembly of the multigene transfer vector is done *in vitro* by fusion of acceptor (pFL) and donors (pUCDM, pSPL) plasmids derivatives using Cre recombinase (**Figure 7**).

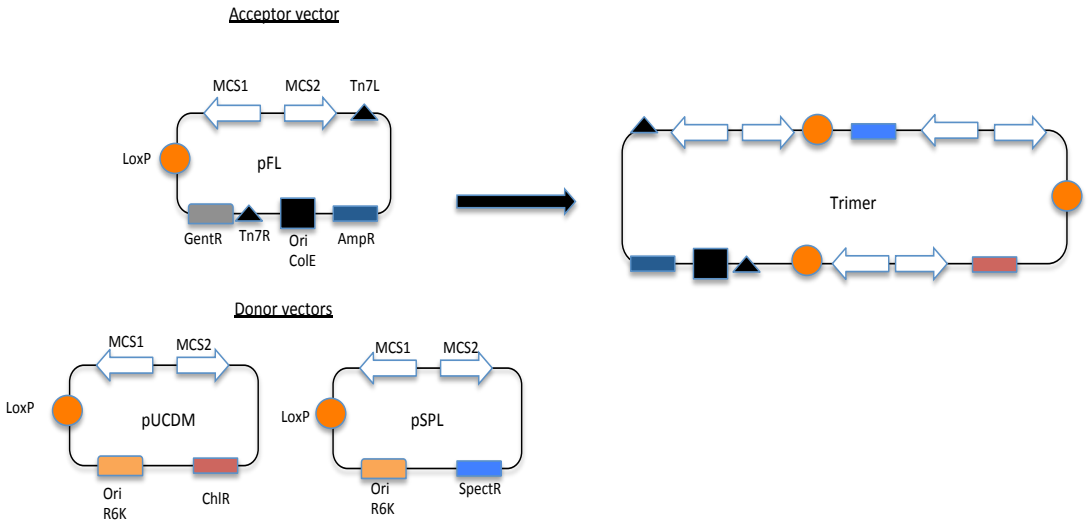



**Figure 7. Scheme of the multigene transfer vectors and its assembly by Cre recombinase**. Acceptor and donor plasmids contain the loxP imperfect inverted repeat. Derivatives of one acceptor plasmid can be fused to one or two donor plasmids generating multigene plasmid. Expression cassettes in between the Tn7 transposition sequences are integrated into the MultiBac baculovirus genome (Adapted from Fitzgerald Daniel J. et al., 2006).

### 3.2.1. Cloning hMCM2-7 in MultiBac Expression System

The human Mcm 2-7 genes were cloned in tandem using the restriccion sites available in the Multiple Cloning Site (MCS) of the corresponding vector to generate the bacmid (**Table 2**). The Mcm7 subunit contained 6x Histidine tag followed by a Strep-tag at the C-terminus end.

| Vector | hMCM subunit | Restriction Site used |
|---|---|---|
| pFL | Mcm6/Mcm7 | EcoR I/Sal I (Mcm6); Xma I/Nhe I (Mcm7) |
| pSpL | Mcm2/Mcm4 | BamH I/Xba I (Mcm2); Xma I /Nhe I (Mcm4) |
| pUCDM | Mcm3/Mcm5 | Xma I /Nhe I (Mcm3); Sal I / Xba I (Mcm5) |

**Table 2. Resume of the restriction sites used for each Mcm subunit**.

After the expression of the hMCM2-7 in insect cells, we obtained very-low-yield of purified complex due to the poor exposure of the tag at the C-terminal end of the Mcm7 subunit used in the first step of purification. Consequently we decided to re-clon the subunits to remove the mcm7's tag by the insertion of a stop codon at the beginning of the tag, and to insert a new tags at the N-terminus end of the Mcm2, HA-tag and Strep-tag, and FLAG-tag, 8x His-tag and TEV cleavage site on Mcm4's N-terminus part.

### 3.2.1.1 hMcm7

We used 50 ng of the pFL6+7 vector as a template to amplify the Mcm7 and insert the stop codon at C-terminus end of the Mcm7, using the Expand Long Template Polymerase (Roche) and the MCM7(fw) and MCM7(rv) primers (**Table 3**). The PCR conditions are described in **Table 4.** The PCR product was digested with Dpn I (NEB) at 37 °C for 2 hours and the mixture was transformed in *E.coli* TOP10 host strain on LB-Ampicilin plates. The different clones were checked by restriction analysis and DNA sequencing using specific primers for the multiple cloning site and internal Mcm7 primers.

### 3.2.1.2 hMcm2 and hMcm4

To insert the corresponding tags in the N- terminal of Mcm2 and Mcm4, we used the pSpL2+4 vector as template to amplify each cDNA. The PCR products were cloned into a pCR-BluntII-TOPO vector (Invitrogen). The resultant TOPO vectors were digested using restriction enzymes to excise the Mcm2 or Mcm4 cDNA and insert them into a pre-digested empty pSpL vector.

We used 30 ng of pSpL2+4 vector and the primers JTM4(fw), MCM4(rv) (**Table 3**) and Vent polymerase (Roche). The PCR conditions are described in **Table 4**. The PCR product was clonned into the pCR-Blunt II-TOPO vectors using the Zero Blunt TOPO PCR Cloning Kit

(Invitrogen) and transformed into *E.coli* TOP10 strain. Positive clones were selected on LB-Kanamycin plates. Positive clones were checked with M13 primers (Invitrogen) to confirm that the FLAG-tag, the 8x His-tag and the TEV cleavage site were correctly inserted in the N-terminal end of Mcm4 subunit. The TOPO+mcm4 vector was digested using Xma I/Nhe I restriction enzymes. The band corresponding to the Mcm4 gene was extracted from the gel using DNA extraction kit (Bio-Rad). Then, an empty vector pSpL was also digested with Xma I/Nhe I present in the MCS2 cloning site and treated with alkaline phosphatese SAP (Roche) for 2 hours at 37 °C . Finally, we inserted the tagged Mcm4 cDNA into the digested pSpL vector using the T4 DNA Ligase incubating the mixture at 16 °C overnight. The ligation mixture was used to transform *E.coli* BW23474 competent cells. The culture was plated on LB-Spectinomycin agar plates for selection and positive clones were analyzed by restriction enzyme digestion and sequencing using MCS2 specific primers (**Table 5**).

To amplify the Mcm2 cDNA, we used 30 ng of pSpL2+4 plasmid as template and Vent polymerase (Roche) together with the primers JTM2(fw) and MCM2(rv) (**Table 3**). PCR conditions described in **Table 4**. This PCR product was inserted into the pCR-Blunt II-TOPO vectors using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen) and trasnformed in *E.coli* TOP10 strain. Positive clones were selected in on LB-Kanamycin agar plates. The positive clones were checked by DNA sequencing using the commercial available M13 primers (Invitrogen). Then, the DNA plasmid was digested with BamHI/Xba I (Roche), loaded into 0.8% agarose gel and the band corresponding to the Mcm2 was extracted from the gel and isolated using DNA extraction kit (Bio-Rad). On the other hand, the pSpL+Mcm4 plasmid was digested with BamH I/Xba I situated at the MCS1 and treated with alkaline phosphatese SAP (Roche) for 2 hours at 37 °C . We ligated the Mcm2 subunit into the pSpL+Mcm4 vector using the T4 DNA Ligase (NEB) at 16 °C overnight. The resulting vector was transformed in the chemical competent *E.coli* BW23474 cells and plated on LB-Spectinomycin agar plates. The clones were analyzed by restriction enzymes and by sequencing using primers designed for each MCS (**Table 5**).

| Primers | Sequence |
|---|---|
| JTM2(fw) | **5'**GGATCCACCATGGGATGGAGCCACCCGCAGTTCGAAAAAGGAAGTGGAGA GAATCTTTATTTTCGGGCGCGGAATCATCGGAATCCTTCACC **3'** |
| MCM2(rv) | **5'** GTGGTATGGCTGATTATGATC **3'** |
| JTM4(fw) | **5'**CCCGGGACCATGGGTGATTACAAGGATGACGACGATAAGCATCATCATCATCAT CATCATCATGAAGTGGAGAGAATCTTTATTTTCAGGGCTCGTCCCCGGCGTCG **3'** |
| MCM4(rv) | 5' GTATTGTCTCCTTCCG 3' |
| MCM7(fw) | **5'**GGATCACTTTTGTCTGAGAAAACCTTTAC **3'** |
| MCM7(rv) | **5'**GTAAAGGTTTTCTCAGACAAAAGTGATCC **3'** |

**Table 3. Primers used for amplification and cloning of the human mcm subunits 2, 4 and 7.**

| Gene | PCR conditions |
|---|---|
| Mcm7 | 5 min 94 °C<br>30 sec 94 °C<br>1 min 52°C<br>8 min 72 °C<br>15 min 72 °C<br>} 20 cycles |
| Mcm4 | 5 min 94 °C<br>1 min 94 °C<br>1 min 46 °C<br>3 min 72 °C<br>5 min 72 °C<br>} 20 cycles |
| Mcm2 | 5 min 94 °C<br>1 min 94 °C<br>1 min 55 °C<br>3 min 72 °C<br>5 min 72 °C<br>} 20 cycles |

**Table 4. PCR conditions for the hMcm7, hMcm4 and hMcm2 genes.**

| Primers | Sequence |
|---|---|
| MCS1(fw) | **5'** TTTTGTAATAAAAAAACCTAT **3'** |
| MCS1(rv) | **5'** TTTAAAGCAAGTAAAACCTCTAC **3'** |
| MCS2(fw) | **5'** TTTATTGCCGTCATAGCGCGGGTT **3'** |
| MCS2(rv) | **5'** TTATCAAATCATTTGTATATTAATT **3'** |

**Table 5. Primers corresponding to the Multiple Cloning Sites 1 and 2 valid for all MultiBac Expression System vectors.**

All Mcm subunits were also sequenced with internal primers to confirm that no mutation has occurred during the cloning (see **Table 6**).

| Primer | Sequence | Primer | Sequence |
|---|---|---|---|
| Mcm2(1) | CCTGGAGGTGGTACTGG | Mcm3(3) | ACCGTTACAGAGCACC |
| Mcm2(2) | CCAGGCCTCTCTTGATGT | Mcm5(1) | CCATCAAGAAGTTTGGCC |
| Mcm2(3) | CCAGGCCTCTCTTGATGT | Mcm5(2) | GACGACCTTGTCACACA |
| Mcm2(4) | CTTGTCCTGGTCCATCTG | Mcm5(3) | GGCGATAGAGATGGTCTG |
| Mcm4(1) | GGGTAACGGTCAAAGAAG | Mcm6(1) | GTTTCTGCACGTGCTCC |
| Mcm4(2) | CTTCTTTGACCGTTACCC | Mcm6(2) | GGGACACTGATTGTTGTGC |
| Mcm4(3) | ATGAGAGGCTTGCTTCA | Mcm6(3) | TCAGCTCCCATCATGTCC |
| Mcm4(4) | AGGACTACATTGCCTACG | Mcm7(1) | TTCTTGCCAATCCTGCG |
| Mcm3(1) | CCCTCTTTCCAGGAAGG | Mcm7(2) | GGGCAATCCTTGTGTTCT |
| Mcm3(2) | GGTGGATAAAGCGAAGCC | Mcm7(3) | CCAGAGGTTCAAACTGGG |

**Table 6. Specific primers designed to verify whole mcm subunit genes.**

### 3.2.2. Recombination of vectors for MultiBac Expression System

The Cre recombinase catalyzes the site-specific recombination of DNA between *loxP* sites. The reaction was performed with 500 ng of acceptor vector (pFL6+7) and 500 ng of one of the donor vectors (pSPL2+4). 5ul Cre recombinase enzyme (NEB) were mixed with the plasmids for 30 minutes at 37 ºC and the sample was transformed in competent *E.coli* TOP10 cells. Positive clones were selected on LB-Ampicilin+Spectinomycin plates. The plasmid DNA of the positive clones, pFL6/7+pSPL2/4, was extracted using QIAprep Spin Miniprep kit (Quiagen) and analyzed using the corresponding restriction enzymes. 500 ng of the fused vector was used for the next Cre recombination reaction together with 500 ng of a second donor vector (pUCDM3/5), in the same conditions as the previous reaction and transformed in competent *E. coli* TOP10 cells and plated on LB-Ampicilin+Spectinomycin+Chloranphenicol agar plates. The plasmid DNA of positive clones was extracted with QIAprep Spin Miniprep kit and analyzed with restriction enzymes.

The multigene transfer vector pFL6/7+pSPL2/4+pUCDM3/5, containing all six hMcm genes, was then integrated into the MultiBac baculovirus genome directly by Tn7 transposition in DH10MultiBac cells (Fitzgerald Daniel J. et al., 2006). 10 ng of DNA were incubated with 50 ul DH10Bac electro competent cells for 15min in ice, transferred into a electroporation cuvette

(BioRad, Gene Pulser Cuvette 0.1 cm) and electroporated at 200 Ohms, 25 uF and 2.0 kVolts. The electroporated cells were mixed with 950ul 2x TYmedia, tranferred into a 14ml tube and incubated with shaking overnight at 37 °C at 220 rpm. 300 ul of the culture were plated onto Low Salt plates containing Gentamycin 10 ug/ml, Ampicillin 100 ug/ml, Spectinomycin 50 ug/ml, Chloramphenicol 30 ug/ml and Tetracyclin 10 ug/ml, 0.5 mM IPTG and Blue-Gal 500 ug/ml. Plates were incubated at 37 °C for 48 hours until a difference was observed between blue and white clones (**Table 7**).

| *Vector* | *Genes and restriction sites* | *Tags* | *Recombination sequences* | *Host strain* | *Antibiotic* |
|---|---|---|---|---|---|
| pFL6+7 | Mcm6(EcoRI/SalI) Mcm7(XmaI/NheI) | | Tn7L, Tn7R, loxP | TOP10 | Ampicillin (100 ug/ml) |
| pSpL2+4 | Mcm2(BamHI/XbaI) Mcm4(XmaI/NheI) | HA-Strep-Mcm2 FLAG-8xHis-TEV-Mcm4 | loxP | BW23474 | Spectinomycin (50 ug/ml) |
| pUCDM5+3 | Mcm5(SalI/XbaI) Mcm3(XmaI/NheI) | | loxP | BW23474 | Chloramphenicol (30 ug/ml) |
| Multibac bacmid derivative of pFL | | | Tn7L, Tn7R, loxP | DH10MultiBac | Tetracyclin (10 ug/ml) Gentamycin (10 ug/ml) Ampicillin (100 ug/ml) |

**Table 7. Resume of clones used for expression of the multiprotein human complex MCM2-7**

### 3.2.3. Bacmid DNA preparation

DH10MultiBac hMCM2-7 positive clones (clear white colony) were grown in 10 ml 2x TY containing Ampicillin 100 ug/ml, Spectinomycin 50 ug/ml, Chloramphenicol 30 ug/ml,

Tetracyclin 10 ug/ml and Gentamycin 10 ug/ml for 24 hours. After collecting the cells for centrifugation, cells were disrupted by alkaline lysis using solutions I, II, and III of the QIAprep spin miniprep kit following the protocol provided by the manufacturer (Quiagen). The DNA was precipitated using isopropanol, and the pellet was washed two times with 70% ethanol. Finally, the bacmid DNA was resuspended in 30 ul of filter-sterilized distilled water to a final concentration of 4.6 ng/ul.

### 3.2.4. Insect cells transfection and hMCM2-7 baculovirus generation

*Spodoptea frugiperda* Sf21 cells were grown in SF-900 II SFM (GIBCO) media containing 1% Pluronic and 0.1% Gentamycin, from now on this media will be called *complete SF-900 II SFM*.

In a sterile hood, 2 ml of $0.5x10^6$ freshly diluted Sf21 cells per milliliter were seeded in four wells of the 6-well tissue plate and fixed for an hour at 27 °C . Meanwhile, the composite bacmid (30 ul at 4.6 ng/ul) was mixed with 200 ul complete SF-900 II SFM and 10 ul FuGENE (Promega) in an 1.5 ml eppendorf tube and incubated for 30 min at 27 °C . 100 ul of this mixture was added to each of the two wells prepared for each composite bacmid to perform the transfection. 2 ml at $0.5x10^6$ cells/ml non-infected control cells, and 2 ml of complete SF-900 SFM were prepared on the other two free wells of the 6-well plate to check whether contamination has occur during the transfection. The 6-well plate was placed into a box containing wet paper to keep high humidity and incubated at 27 °C for 72 hours before collecting the first V0. The first V0 was transferred into a 15 ml sterile tube and stored in dark at 4 °C sealed with Parafilm. Once collected, another 2 ml of complete SF-900 II SFM were added onto the cells, the second V0 was collected 72 hours latter and stored at 4 °C . The infected cells and the control cells were resuspended in 1 ml complete SF-900 II SFM each, transferred into a 1.5 ml eppendorf tube and pelleted for 5 minutes at 1000 rpm in a benchtop centrifuge. The supernatant was discarded and the pellet was stored at -20 °C . To analyze the protein expression, the pellets were resuspended in 500 ul 1x PBS (10x PBS contains: 76.8 mM Na2HPO4, 76.8 mM NaH2PO4, 1.54 mM NaCl, pH to 7.2 adjusted with HCl) and 500 ul 6x Loading Dye (125 mM Bis-Tris pH:6.8, 20% v/v glycerol, 4% w/v Sodium Dodecyl Sulfate, 10% w/ β-mercaptoethanol, 0.4 mg/ml bromophenol blue), and sonicated for 30 seconds to lyse them. The protein expression was analyzed in a SDS-PAGE gel and Western Blots against each Mcm subunit.

To amplify the hMCM2-7 virus, 25 ml of freshly diluted Sf21 cells at 0.5x10$^6$ cells/ml were prepared in a 125 ml shaker flask and the 2 ml of the first V0 were added into the culture. At that point 1x10$^6$ cells were collected and used as a control cells before the complex over-expression. Sf21 cells were incubated at 27 ºC with shaking at 90 rpm. Concentration of cells was measured 24 hours after infection, if density was over 1x10$^6$ cells/ml, the culture was diluted to 0.5x10$^6$ cells/ml otherwise the cells were not diluted. 1x10$^6$ cells were taken at that time to confirm protein over-expression. When concentration of the cells was lower than 1x10$^6$ cells/ml and cells were double sized than the non-infected cells, that time was considered as Day after Proliferation Arrest (DPA), and 1x10$^6$ cells were retrieved to confirm protein over-expression. The DPA ocured 48-72 hours after infection. Next day after DPA (DPA+24), the culture was centrifuged in 50 ml Falcon tubes for 5 min at 800 x *g*. The supernatant (V1 virus) was collected and stored in a fresh sterile 50 ml Falcon tube at 4 °C in the dark.

The cells were pelleted in each step and used to measure the over-expression of the human MCM2-7 complex in a SDS-PAGE and by Western Blots against each Mcm subunit.

For creation of the second generation of the virus (V2), 5-10 ml of V1, depending of its titration, were added to 400 ml of previously diluted Sf21 cells at a density of 0.5x10$^6$ cells/ml in a 2-liter Erlenmeyer flask and incubated at 27 °C and 90 rpm. Cells were counted every 24 hours and diluted when density was over 1x10$^6$ cells/ml. When the cells were completely infected, proliferation arrest was observed. When infected the cells are swelled two-fold compared with non-infected cells. The V2 was collected 48 hours after DPA by centrifuging the cell suspension in 50 ml Falcon tubes for 5 min at 800 x *g*. The supernatant was poured directly into a sterile 50 ml Falcon tube and stored at 4 °C or at -80 ºC protected from light.

### 3.2.5. Virus titration and small scale expression protocol

Virus titration is a process in which serial dilutions of virus in the cell culture are done to define the lowest concentration of virus required for infecting a certain amount of cells. For that, we prepared several flasks with 25 ml of Sf21 culture at 0.5x10$^6$ cells/ml and we added different volumes of hMCM2-7 V1 100 ul, 250 ul, 500 ul, 1 ml and 2 ml. We counted the cells every 24 hours and collect 1x10$^6$ cells. In the case that cells stopped their proliferation

immediately, the volume of V1 to use should be lower; if cells were still dividing 48 hours after infection, higher concentration of V1 should be used. Ideally, the volume of virus to use, should provoke DPA 24 hours after infection and cells should be collected at DPA+48. Over-expression of the complex was checked by SDS-PAGE and Western Blots against each of the Mcm subunits. To titer the V2, the same procedure was followed.

## 3.3. Cell culture

### 3.3.1 Sf21 cells

The Sf21 cells were stored in a liquid nitrogen tank in a sterile cryovial containing 1 milliliter of cells at $1x10^7$ cells/ml and resuspended in 60 % SF-900 SFM, 30 % FBS and 10 % DMSO. To thaw the cells, the cryovial was placed in a water bath at 37 °C and directly transferred to a 25 $cm^2$ flask containing 4 ml of complete SF-900 SFM. The flask was then transferred to the 27 °C incubator allowing cells to attach to the flask bottom for 45 minutes. Then the medium was removed and other 5 ml of fresh medium were added onto the cell layer. Once the cells have grown up to 80 % or 90 % confluence (percentage of surface occupied by cells), we moved them into 150 $cm^2$ flask, containing 20 ml of complete SF-900 SFM, where they were grown as a monolayer. The culture medium was removed every 2 or 3 days and substituted by 25 ml of a fresh medium.

When cells reached 90 % or 100 % confluence in T150 flasks, the cells were resuspended in the 25 ml medium contained in th eT150 flask. 1 milliliter of the resuspended cells was added to a new T150 flask containing 24 ml of fresh complete SF-900 II SFM. Each time we have transferred the cells into a new T150 flask, it was considered as one passage of cells. Cells were diluted up to 30 passages when grown in a monolayer.

For Sf21 cell suspension culture, we used Erlenmeyer shaker flasks from 125 ml up to 2-liters always filled up to 20 % of the total volume for proper aeration and optimal cell growth. Cells have been grown up to a maximal density of $4x10^6$ cells/ml and then were diluted to $0.3x10^6$ cells/ml. Each cell dilution was considerred as one cell passage. Cells were grown not more than 28-30 passages.

For a large scale protein expression, 2-liter Erlenmeyer flasks filled up to 25 % of the total volume were used, also WAVE bioreactor system(GE Healthcare) with 10-liters bags filled up

to 5 liters of cell culture. The ratio of hMCM2-7 baculovirus was 50 ml of V2 for 5 liters at 0.5x10$^6$ cells/ml or 50 ml of V2 for 2 liters at 1x10$^6$ cells/ml.

Cells were collected at DPA+48, and centrifuge in 1-liter centrifugation flasks at 800 x *g* for 20 minutes at 4 °C. The supernatant was discarded and the pellet was transfered in a sterile 50 ml Falcon tube, quickly frozen in liquid nitrogen and stored at -80 °C until needed. The volume of frozen pellet corresponding to 1 liter culture infected at 1x10$^6$ cells/ml is between 15 to 20 ml.

### 3.3.2 HeLa cells

The HeLa cells were stored in a liquid nitrogen tank in a sterile cryovial. After thawing the cells in water bath at 37 °C , the cells were resuspended in 4ml DMEM and transferred into 25 cm² flask. These cells were grown as monolayer. Once HeLa cells reached the 70% confluence, they were resuspended in 20ml pre-warmed DMEM and transferred into a 150 cm² flask.

## 3.4. Isolation of hMCM2-7 protein complex

The pellet of infected Sf21 cells was resuspended in a pre-cooled lysis buffer (50 mM NaH2PO4, 300 mM potassium acetate, 10 mM imidazole, 0.02% Tween, 0.4 mM PMSF, 5 mM magnesium chloride and 10% glycerol, pH: 8) with 200 ul of DNAse I (5000 units/ml) and incubated for 10 minutes in ice. The cells were later sonicated at 4 °C for 10 minutes at 0.5 seconds pulses and the cell extract was clarify by centrifugation for 45 minutes at 40000 rpm at 4 °C using a 45Ti rotor (Beckman). The supernatant was incubated with pre-equillibrated Ni-NTA resin (Quiagen) using Wash Buffer (50 mM $NaH_2PO_4$, 300 mM Potassium Acetate, 20 mM imidazole, 0.4 mM PMSF and 10% glycerol, pH:8) in batch for 1 hour at 4 °C . Later on, the resin was packed in a column using a peristaltic pump and connected in to a FPLC Äkta purification system. Proteins that did not bind to the resin were eliminated by washing the resin with 10 volumes of wash buffer and a non-specific proteins bound to the resin were removed by serial of washes at 3% and 7% of Elution Buffer (the wash buffer supplemented with 500 mM imidazole), corresponding to 35 mM imidazole and 55 mM imidazole respectively. Finally, the elution of the hMCM2-7 complex was performed by linear gradient

of 20 volumes of resin from 55 mM to 500 mM imidazole using elution buffer. The fractions containing the complex were pooled together.

In the second purification step we used an ionic exchange chromatography column, HiTrap SP FF 5 ml (GE Healthcare). The protein eluted from the Ni-NTA was diluted in Dilution Buffer (25 mM Tris pH:7.5, 1 mM EDTA, 0.4 mM PMSF and 10% glycerol), for decreasing the amount of salt, to a final salt concentration of 75 mM potassium acetate. The elution fraction was passed through the HiTrap SP FF column equilibrated with Heparin Wash Buffer (25 mM Tris pH:7.5, 50 mM potassium acetate, 1 mM EDTA, 0.4 mM PMSF and 10% glycerol) using a peristaltic pump with a flow rate of 3 ml/min. In this buffer conditios the flow-through fraction contains the hMCM2-7 complex.

As a third purification step, the flow-through from SP FF column was loaded into a HiTrap Heparin HP 5 ml column (GE Healthcare) at 0.5 ml/min flow rate, previously equilibrated with Heparin wash buffer, and washed with 50 ml wash buffer to remove any protein not bound to the resin. The elution is done by linear gradient of 20 resin volumes of Heparin elution buffer (wash buffer supplemented with 1.5 M potassium acetate) from 50 mM to 1.5 M potassium acetate.

As a last step of purification, the elution fractions were concentrated by using a Vivaspin20 filtration unit (Sartorius Stedim Biotech) with a cut-off of 50000 MWCO, PES membrane and injected into a gel filtration column Superdex 200 16/60 (GE Healthcare) equilibrated with the Superdex Buffer (20 mM Bis-Tris pH:7, 150 mM potasium glutamate, 5 mM magnesium chloride, 10% glycerol). The elution fractions, containig the hMCM2-7 heterohexamer, were pooled together and concentrated in Vivaspin20 filtration unit with addition of 2 mM ATPgammaS. The fractions containing the heterohexameric complex were analyzed in 12% SDS-PAGE or AnyKD gels (Bio-Rad).

## 3.5. Biochemical analysis of the hMCM2-7

### 3.5.1. DNA binding assay

The hMCM complex bound to DNA was purified by binding specific biotin-labeled oligonucleotides combined with streptavidin-covered magnetic beads (Dynabeads M-280

Streptavidin, Invitrogen). We designed three different structures of DNA labeled with biotin (**Table 8**). The EcoR I restriction site was located 25 base-pairs downstream from the biotin.

| *DNA probe* | *Sequence* |
|---|---|
| Single strand(fw) (overhangDNA) | **5'**ATCTATACATACACGTGCCTTCGATAATGGGAATTCAAACAGTAGCATGCATACACC GTA-**3'** |
| Single strand(rv) (overhangDNA) | **Btn5'**TACGGTGTATGCATGCTACTGTTTGAATTCCCATTATCGAAGGCACGTGTATGTA TAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-**3'** |
| Bubble(fw) | **Btn5'**ATCTATACATACAGTTGCGTTCGATAGTGGGAATTCAAAGAGTAGGATCCATAC ACCGTCAGTATAAGCATCTAAGTTGAACGTCTAGACAGTCAGCGATTAAAGTTGTATA CACG-**3'** |
| Bubble(rv) | **5'**CGTGTATACAACTTTTTAGCGACTGACAGATCTGCAAGTTGAATCTACGAATATG ACTGCCACATTGGATCCTACTCTTTGAATTCCCACTATCGAACGCAACTGTATGTAT AGAT-**3'** |
| Fork(fw) | **Btn5'**ATCTATACATACAGTTGCGTTCGATAGTGGGAATTCAAAGAGTAGGATCCATAC ACCGTCAGTATAAGCATCTAAGTTGAACGTCTAGACAGTCAGCGATT-**3'** |
| Fork(rv) | **5'**TTAGCGACTGACAGATCTGCAAGTTGAATCTACGAATATGACTGCCACATTGGA TCCTACTCTTTGAATTCCCACTATCGAACGCAACTGTATGTATAGAT-**3'** |

**Table 8. Oligonucleotides used for DNA binding assays.**

The corresponding pairs of oligonucleotides were hybridated by mixing them at final concentration of 50 mM and by addindg 15 mM sodium chloride to the reaction. The mixture was incubated at 94 °C for 4 minutes and then cooled down slowly to the room temperature. The hybridazation of oligos was checked in 2% agarose gel.

The sterpatividin-coated magnetic beads were washed twice with water and pre-equilibrated with Superdex Buffer (20 mM Bis-Tris pH:7, 150 mM sodium glutamate, 5 mM magnesium chloride and 10% glycerol). The 20 ul beads were later resuspended in 100 ul of Superdex Buffer and 3 ul hybridated DNA at final concentration of 15 mM. The binding reaction of DNA onto the beads was done at room temperature for one hour with agitation. 100 ul of hMCM2-7 complex, at 0.1 mg/ml, were added to the DNA-beads in presence of 2 mM ATPγS and incubated for 3 hours at 4 °C with shacking. To remove the hMCM2-7 complex not bound to DNA, the beads were fixed with magnet and the supernatant was removed. Later, the beads were washed three times with superdex buffer and resuspended

in 30 ul the buffer together with 1 ul Eco RI (Roche) and incubated over-night at 4 °C. After that, the beads were fixed with the magnet, and the elution fraction containing the complex with DNA was collected (**Figure 8**).

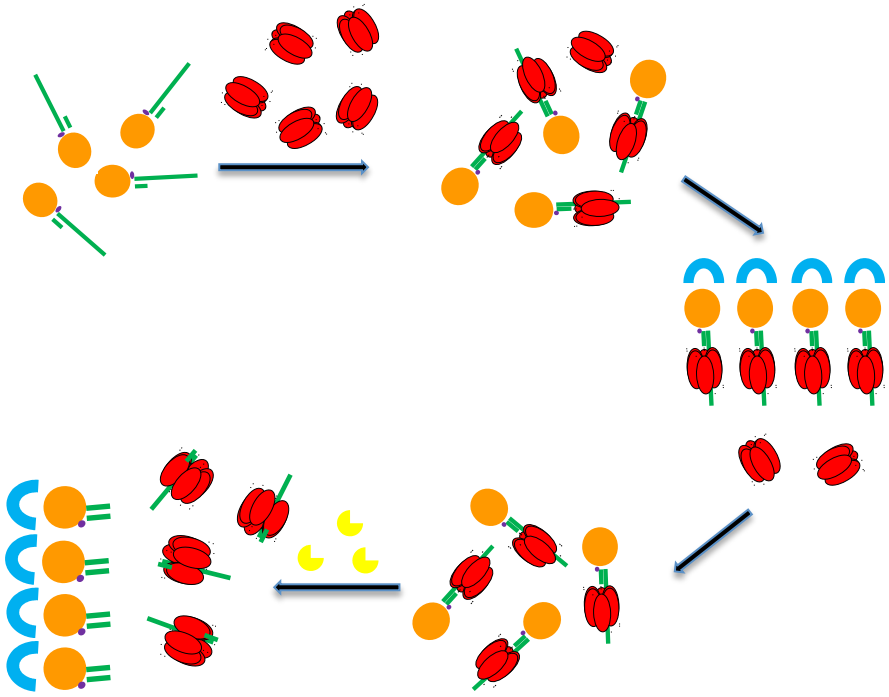


**Figure 8. Representation of purification of the hMCM2-7+DNA.** Firstlly, the DNA with Biotin was bound to the streptavidin magnetic beads, then the hexamer hMCM2-7 was incubated with the DNA. After incubation, the beads together with the DNA and the hMCM2-7 were fixed using DynaMag™-2 Magnet" (Invitrogen) and washed with buffer, eliminating the complex not bound to the DNA. Finally, the complex with DNA was eluted from the beads using EcoR I restriction enzyme.

### 3.5.2. ATPse assay

The ATPase assay is based on a reaction in which the regeneration of hydrolyzed ATP is coupled to the oxidation of NADH. In this regeneration system, the Pyruvate Kinase (PK) converts one molecule of phosphoenolpyruvate (PEP) to pyruvate when the ADP is converted back to the ATP, meanwhile ATP is hydrolyzed by the hMCM2-7. The pyruvate is subsequently converted to lactate by L-lactate Dehydrogenase (LDH) resulting in the oxidation of one NADH molecule (Panuska & Goldthwait, 1980). The assay measures the rate of NADH absorbance decrease at 340 nm, which is proportional to the rate of steady-state ATP hydrolysis. The constant regeneration of ATP allows monitoring the ATP hydrolysis rate over the entire course of the assay (**Figure 9**).

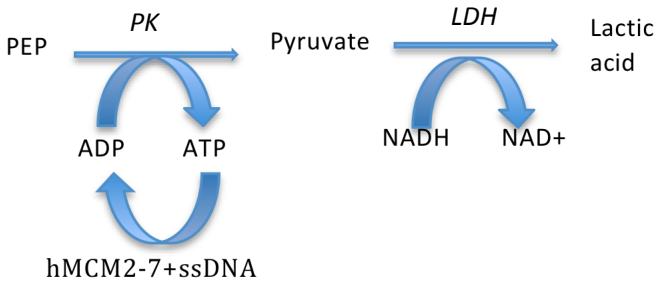



**Figure 9. The pyruvate kinase/lactate dehydrogenase coupled assay used for the measurement of the ATPase activity of hMCM2-7**

The assay was carried in 20 ul final volume, in ATPase buffer (10 mM Bis-Tris pH:7, 10 mM $MgCl_2$, 0.1 mg/ml BSA) with 5 mM PEP, 2 mM NADH, 1 mM ATP, 75 units/ml of PK and 125 units/ml LDH (Sigma) incubated at 30 °C . The absorbance was measured in a NanoDrop spectrophotometer (ND-1000) using aliquots of 2ul measured each 3 min during 15min. The protein concentration used was 1 uM, and 5 nM of the overhangDNA probe.

### 3.5.3. Helicase assay

We used a 3' overhang DNA structure to determine the helicase activity of the hMCM2-7 complex. *Oligo B* was labeled with [$\gamma^{32}$P] ATP at the 5' end using T4 polynucleotide kinase, and then hybridized with *Oligo A* overnight in presence of 150 mM sodium chloride (see **Table 9**). The helicase activity was performed *in vitro* by incubating for 1 hour at 37 °C the 0.75 uM or 1.5 uM of hMCM2-7 complex in presence of 0.5 nM DNA in Helicase Buffer (20 mM BisTris pH:7.5, 150 mM Potassium Glutamate, 10 mM Magnesium Acetate, 10% glycerol, 0.1 mg/ml BSA, 2 mM DTT), supplemented with 5 mM ATP or ATPγS , in a final volume of 20 ul,. The reaction was stopped using 5x loading dye supplemented with 100 mM EDTA and 0.2% SDS (10 mM Tris-HCl pH=7,6, 60 % glycerol, 0,03 % bromophenol and 0,03 % xylen cyanol). The samples were solved by electrophoresis in 15% non-denaturing acrylamide gels using TBE buffer and at 13.3 V/cm for 3 hours. The reaction without protein, used as a control, was boiled at 95 °C for 10 minutes before being loaded into the gel. After the electrophoresis, the gels were dried on 3MM Whatman paper and exposed on a film at -80 °C over-night to increase the signal of the labeled DNA.

| *DNA probe* | *Sequence* |
|---|---|
| Oligo A | 5'ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCG-3' |
| Oligo B (3' overhang) | 5'-GTTGTAAAACGACGGCCAGT-3' |

**Table 9. Sequences of the oligonucleotides used for the helicase assay**. The Oligo B was labeled radiactivelly. Oligo A was hibridazed with Oligo B to generate the 3' overhang structure with a 3' free end.

## 3.6. Structural analysis of the hMCM2-7

### 3.6.1. Electron microscopy

#### 3.6.1.1. Sample preparation and dataset generation

For negative staining, we have used 400 mesh copper grids covered with carbon layer (Electron Microscopy Science).A glow discharge was applied for 30 seconds to make the grids more hydrophilic and to improve the binding of the sample. 4 ul of hMCM2-7 complex dilututed to 0.1 mg/ml were incubated for 1:30 minutes on the grid at room temperature. The excess of protein not bound onto the grid was eliminated by blotting with filter paper. The grid were washed twice by placing it at the top of two distinct MiliQ water drops and the excess of water was eliminated again by blotting the grid with filter paper. Finally, the grids were stained with 2% uranyl acetate, or 1% uranyl formate, and incubated for another 1:30 minutes. The excess of the staining agent was removed with filter paper and the grid was air-dried at room temperature for at least 10 minutes.

The grids were observed using a Tecnai 12 transmission EM (FEI) with Lanthanum hexaboride cathode operated at 120 keV. The images were recorded under low-dose conditions with an electron dose of 8-10 $e^-/Å^2$ at 61,320x nominal magnification on a 4k x 4k TVIPS TemCam-F416 CMOS camera (TVIP GmbH) resulting in a final pixel size at the specimen level of 2.5 Å. Several hundreds of images were obtained for each experiment. The contrast transfer function (CTF) was determined on micrographs with ctffind3 (Mindell & Grigorieff, 2003) and its effects in the images were corrected with bctf from bsoft program (http://lsbr.niams.nih.gov/bsoft/programs/bctf.html).

Initially, the particle picking was performed manually using boxer (EMAN1.9) (Ludtke et al., 1999) or semi-automatically with e2boxer.py implemented in EMAN2 (Tang G et al JSB 2007). A total of 21434 particles were picked for apo hMCM2-7, 10370 for ADP-bound $hMCM_{2-7}$, 20394 for ATPγS-bound $hMCM_{2-7}$ and 18268 particles for hMCM2-7–ATPγS–overhangDNA complex. A 128x128 pixel size window was used for all datasets. The particles were centered and normalized using proc2d (EMAN) before its initial classification using EMAN and Xmipp (Sorzano et al., 2004).

### 3.6.1.2. Two-dimensional images analysis and 3D reconstruction

For 2D classification the images were down sampled by a factor of 2. Reference-free 2D classification and averaging of the raw datasets were carried out by refine2d.py implemented in EMAN and CL2D from the Xmipp package (Scheres et al., 2005). A large number of class averages (up to 1000) were produced with around 20 particles in each class.

For both 3D structures, ATPγS-bound $hMCM_{2-7}$ and hMCM2-7–ATPγS–overhangDNA, we used 2D averages as input into the program e2initialmodel.py implemented in EMAN2, to produce eight different starting models. These models were carefully inspected for consistency between original reference-free 2D averages and the models re-projections. For each 3D refinement three different models were selected, low-pass filtered to minimize model bias and used as starting models. We have also used a low-pass filtered "notched" (EMD-1834) and lock-washer (EMD-1835) *D. melanogaster* MCM-2-7 structures. 3D volumes were calculated using an iterative projection-matching approach using libraries from EMAN1.9 and Xmipp. The resolution of both structures was estimated as 28 Å for hMCM2-7–ATPγS and 24 Å for hMCM2-7 –ATPγS–overhangDNA using Fourier Shell Correlation, with 0.5 cut-off criterion. The resulting volumes were filtered in Spider (http://spider.wadsworth.org/spider_doc/spider/docs/spider.html) to calculated resolution using Butter low-pass filtering.

The hand of the three-dimensional reconstruction was decided by comparison of our structures with available crystal structure of a near full length monomer of archaeal MCM (PDB ID code 3F9V). The docking of atomic coordinates was done using UCSF Chimera http://www.cgl.ucsf.edu/chimera/.

### 3.6.2. Cryo-electron microscopy

The cryo-electron microscopy was done at the EMBL Heidelberg in collaboration with the group of Christiane Schaffitzel from the EMBL Grenoble.

The hMCM2-7-DNA complex obtained upon elution with EcoRI and in presence of ATPγS, was concentarted twenty times and then diluted 1:5 in a Superdex buffer without glycerol to decrease its concentration to nearly 2%. The diluted sample was loaded on a Quantifoil 2/2 cryo-grid cover with thin carbon layer and incubated for 4 minutes at room temperature. After 3 seconds blotting for both sides in a Vitrobot System (FEI), the sample was vitrified in ethane near liquid nitrogen temperature.

The grids were observed using a CM 120 Biotwin transmission EM (FEI) operated at 120 keV. The images were recorded with an electron dose of $15e^{-}/Å^{2}$ at 65,217x nominal magnification on a 4kx4k TVIPS TemCam-F415A-CS-HS-4 camera (TVIP GmbH) resulting in a final pixel size at the specimen level of 2.3 Å. 200 images were selected manually for the experiment. The contrast transfer function (CTF) was determined for each micrographs using ctffind3 (Mindell & Grigorieff, 2003) and its effects in the images were corrected with bctf from bsoft program (http://lsbr.niams.nih.gov/bsoft/programs/bctf.html).

Other set of images were recorded with a POLARA working at 100 kV, however due to the low quality of the images we didn't process the particles collected.

### 3.6.3. Crystallography

The initial trials for the crystallization of the hMCM2-7 were carried out with the Cartesian Mycrosys SQ robot (Genomic Solutions) using the sitting drop technique in a 96-well Grainer plate mixing 0.1 ul of protein and 0.1 ul of reservoir buffer and a total volume of 70 ul of reservoir. We set up different commercial screenings, as ProComplex (Quiagen), Wizard1/2 (Biosystems) and CrystalScreen1/2 (Hampton) making almost 500 different conditions. After sitting the drops, the plates were sealed and kept at 4 °C. During the crystallization trials different protein concentrations and different glycerol concentrations were tried.

## 3.7. CMG reconstitution

### 3.7.1HeLa cell line synchronization

#### 3.7.1.1. G1/S arrest

The exponentially growing HeLa cells were platted at 60% to 70% confluence in 20 ml DMEM-10/thymidine (2.5 mM) and incubated 20 hours at 37 °C . The medium containing thymidin was removed and the dishes were rinsed twice with 10 ml prewarmed DMEM, to prevent the carryover of thymidine. Cells were trypsinized using 1 x trypsin and resuspended in 10 ml DMEM-10. The cell suspension was centrifuged for 5 min at 500 x g, and the pellet was washed with 1x PBS and centrifuged again 5 min at 500 x g. Some cells were kept to be analyzed in the flow cytometer FACS ( Flow- Activated Cell Sorter). Remaining the pellet was stored at -20 °C .

#### 3.7.1.2. Mitotic arrest

The HeLa cells were grown at 60% to 70% confluence in DMEM-10 media . Add nocodazole to a final concentration of 50 ng/ml and incubate for 18 hours at 37 °C . Centrifuge cells 5 min at 500 x g at room temperature, discard supernatant and resuspend the pellet in 1x PBS and centrifuge again. The pellet was stored at -20 °C . A small part of the pellet was separated and analyzed in FACS.

#### 3.7.1.3. DNA staining with propidium iodide

The pellet used to check the different cell populations in FACS was resuspended in 70% ice-cold EtOH to fix the cells for at least 30 min at 4 °C . Then, the pellet was washed twice with 5 ml 1 x PBS, and pelleted for 5 min at 500 x g discarding the supernatant. To ensure that only DNA is stained, cells were treated with 50 ul Ribonuclease I (100 ug/ml) (Sigma). This pellet was resuspended in 300 ul propidium iodide (Sigma) at 50 ug/ml, lefted overnight at 4 °C or 1 hour at room temperature. Before loading the cells to the flow cytometer, the cell suspension was passed through a 50 um cut off filter.

We used a FACS Canto II (BD-Bioscience) at the Flow Cytometry Unit in the CNIO (Madrid, Spain) to check the efficiency of the cell synchronization at G1/S interphase and mitotic phase. We used the standard 585/42 Blue filter with a voltage of 300 V. To quantify the percentage of cells in each cell cycle phase, we used the markers set within the analyis program.

### 3.7.1.4. Preparation of G1/S and Mitotic cell extracts

For the preparation of the cell extracts used for CMG assemably, the pellet of HeLa cells was resuspended in Superdex Buffer (20 mM Bis-Tris pH:7, 150 mM potasium glutamate, 5 mM magnesium chloride, 10% glycerol) to a final dilution of 6000 cells/ul and sonicated for 30 sec at 4 °C .

## 3.7.2. CMG assembly

The CMG complex was reconstituted by performing the DNA-binding assay for the hMCM2-7. The oligonucleotides used for this experiments are described in **Table 8**.

A 500 ul of Dynabeads M-280 Streptavidin (Invitrogen) are firstly washed and pre-equilibrated with Superdex Buffer (20 mM Bis-Tris pH:7, 150 mM sodium glutamate, 5 mM magnesium chloride and 10% glycerol). A 12 ul of oligonucleotide at final concentration of 15 mM, are bound to the strepatividin beads for 1 hour at room temperature. Beads are washed with Superdex buffer to eliminate the excess of DNA , and 0.1 mg of hMCM2-7 are incubated with the beads bound to DNA for 4 hours at 4 °C by agitation. The complex not bound onto the DNA was removed by three washes with buffer. Then 500 ul of G1/S cell extract or Mitotic cell extract were added to the beads suplemented with 2 mM ATP or 2 mM ATPγS and 40 ug of human Cdc45 (provided by Ines Muñoz from our lab) and 40 ug of hGINS complex (provided by Javier Coloma former PhD in our lab, see Boskovic et al., 2007). The mixture was incubated for 4 hours at 4 °C with agitation. The proteins not bound to the DNA or to the hMCM2-7 are removed by three washes with the buffer. Finally the beads are resuspended in 500 ul of Superdex buffer and the DNA together with the bound proteins is eluted with 5 ul EcoR I (Roche) at 4 °C overnight with agitation.

For a further purification of the CMG complex, the 500 ul elution was incubated with 400 ul Ni-NTA resin (Quiagen) for 4 hours at 4 °C . The unspecific bound proteins were removed by three washes with Superdex buffer supplemented with 20 mM imidazole. Finally the proteins were eluted by resuspending the resin in 200 ul of Superdex buffer supplemented with 500 mM imidazole.

## 3.8. Proteomic analysis

### 3.8.1. Sample preparation

DNA binding proteins and repurified pull-downs (both from replication and mitosis conditions) were subjected to label free proteome analysis. Samples were digested by means of the standard FASP protocol (Wisniewski, J.R. et al., 2009). Briefly, samples were resuspended in UT buffer (8M urea in 100 Mm Tris-HCl, pH=8.01). Proteins were then reduced with 10 mM DTT, alkylated using 50 mM iodoacetamide for 20 min in the dark and the excess of reagents was washed out with UA twice. Proteins were digested with endoproteinase Lys-C (Wako) during 6 hours in a wet chamber (1:50 enzyme to substrate ratio). Finally, samples were diluted in 50 mM ammonium bicarbonate to reduce the urea concentration to 1M and subsequently digested with Trypsin Gold (Promega) overnight at 37 °C. Resulting peptides were further desalted and concentrated using homemade reversed phase micro-columns filled with Poros Oligo R3 beads (Life Technologies). The samples were dried using the Speed-Vac and dissolved in 30 µL of 0.1% formic acid (FA).

### 3.8.2. LC-MS/MS analysis

Desalted peptides were separated by reversed-phase chromatography using a nanoLC Ultra system (Eksigent), directly coupled with a LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific) via nanoelectrospray source (ProxeonBiosystem). Peptides were loaded onto the column (Dr. Maisch, Reprosil-Pur C18 GmbH 3 mm, 200x0.075 mm), with a previous trapping column step (Prot Trap Column 0.3 x 10 mm, ReproSil C18-AQ, 5 µm, 120Å, SGE), during 10 min with a flow rate of 2.5 ml/min of loading buffer (0.1% FA). Elution from the column was made with a 150 min linear gradient (buffer A: 2% ACN, 0.1%FA; buffer B: 100% ACN, 0.1%FA) at 300 nL/min. The peptides were directly electrosprayed into the mass spectrometer using a PicoTip emitter (360/20 OD/ID µm tip ID 10 µm, New Objective) a 1.4 kV spray voltage with a heated capillary temperature of 325°C and S-Lens of 60%. Mass spectra were acquired in a data-dependent manner, with an automatic switch between MS and MS/MS scans using a top 20 method with a threshold signal of 800 counts. MS spectra were acquired with a resolution of 60000 (FWHM) at 400 m/z in the Orbitrap, scanning a mass range between 350 and 1500 *m/z*. Peptide fragmentation was performed using collision induced dissociation (CID/CAD) and fragment ions were detected in the linear ion trap. The normalized collision energy was set to 35%, the Q value to 0.25 and the activation time to 10 ms. The maximum ion injection times for the survey scan and the MS/MS scans

were 500 ms and 100 ms respectively and the ion target values were set to 1E6 and 5000, respectively for each scan mode. Samples were run in duplicates.

### 3.8.3. Data analysis

Raw files were analyzed by Proteome Discoverer (version 1.4.1.14) against a forward-reverse concatenated human database (UniProtKB/Swiss-Prot, 88354 sequences, 26/03/2013 release). Oxidation of methionines was set as variable modification whereas carbamidomethylation of cysteines was considered as fixed modification in the Mascot (Perkins D.N. et al., 1999) search engine (v2.2). Minimal peptide length was set to 6 amino acids, a maximum of two missed-cleavages were allowed and only sequences with IonScore (Mascot) above 20 were considered. Peptides were filtered at 1% FDR. In case that identified peptides were shared by two or more proteins (homologs or isoforms), they were reported by Proteome Discoverer as one protein group.

As stated above, each sample was run in duplicate. Label-free analysis was performed with Proteome Discoverer using the spectral count values determined by the software. Further analysis was done using Excel and MultiExperiment Viewer software (version 4.8.1). The changes in protein abundance between the different samples were measured via a direct comparison of the spectral count values. Protein classification (molecular function, biological process and protein class) was performed by STRING (Franceschini A. et al., 2013) software.
Raw files from the re-purified pull-downs (replication, mitosis) and expressed MCMs in baculovirus were analyzed either by Proteome Discoverer or by MaxQuant (Cox J. et al., 2008) (version 1.4.1.2) (UniProtKB/Swiss-Prot human database, canonical and isoform sequences, 39748 sequences, 01/22/2014 release), including phosphorylation on serine, threonine and tyrosine residues as variable modification. The rest of search parameters were set as above. SequestHT search engine, in conjunction with Percolator provided the list of proteins for Proteome Discoverer. For protein assessment in MaxQuant, at least two unique peptides provided by Andromeda search engine (Cox J. et al., 2011) with a FDR = 1% were required. Results at peptide label were exported to excel for further analysis. Extracted ion Chromatograms (XIC) of the identified phosphopeptides were manually obtained from Xcalibur (version 2.2). In order to normalize the XICs to the total protein amount, iBAQs from MaxQuant were considered.

# 4. Results

## 4.1. Cloning of the human MCM2-7 genes

The human MCM genes were first identified in *S. cerevisiae* (Maine et al. in 1984) in genetic screenings for genes involved in plasmid stability. Six Mcm proteins are conserved from yeast to human, they share nearly 250 amino acids of the AAA+ ATPase domain and form a heterohexameric complex with a molecular weight of 550 kDa (**Figure 10**).

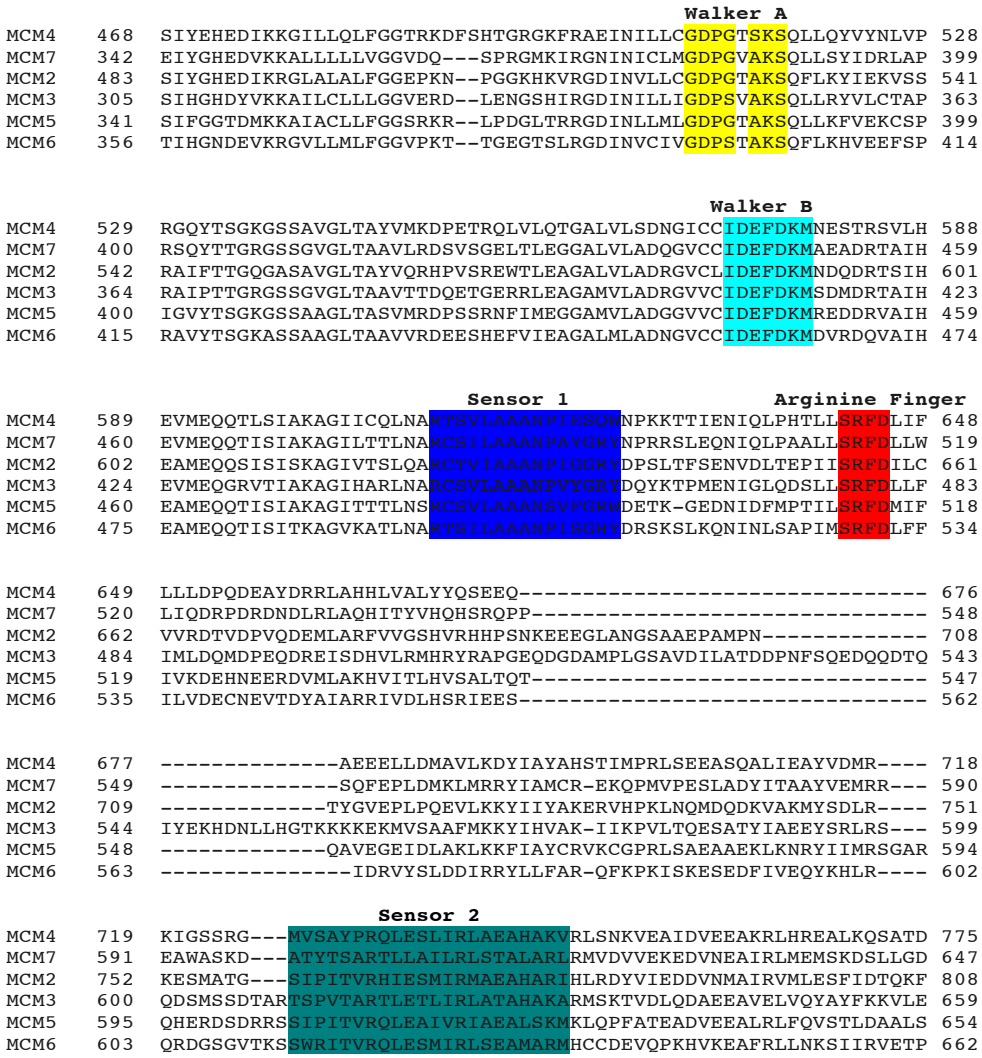



**Figure 10. Shared motifs among the human Mcm subunits**. The non-conserved amino- and carboxy-terminal regions were removed due to spatial constraints. The alignment was done using CLUSTALW. The molecular weights of each human subunit are: Mcm2: 102 kDa; Mcm3: 90 kDa; Mcm4: 96 kDa; Mcm5: 82 kDa; Mcm6: 93 kDa; Mcm7: 81 kDa.

The human Mcm2-7 genes were cloned into the MultiBac System. This insect cell expression system is suitable for multiprotein eukaryotic complex expression. This method is more suitable than yeast or bacterial expression systems for the production of biologically active mammalian proteins. The production of these complexes in insect cells allows protein post-translational modifications (e.g. glycosylation, palmitolation, myristolation, fatty acid acylation,

amidation) that are similar to those of mammalian cells. A conventional approach in baculovirus to produce the six different subunits would need to co-infect insect cells with six different viruses, making this procedure cumbersome in order to maintain simultaneously six viruses with high titer. To solve this problem we decided to use the multibac system to produce all the hMCM2-7 subunits simultaneously from a single multigene baculovirus. Thus, all genes encoding for the hMCM2-7 subunits were combined into a single clone by tandem recombination. The multigene expression cassettes were integrated into the engineered MultiBac baculovirus genome that is characterized by reduced proteolysis and delayed host cell lysis, thus improving the quality of the proteins produced REF. The recombinant MultiBac baculovirus was used to infect Sf21 insect cell cultures to produce the human MCM2-7 complex.

We used simultaneously bioinformatics tools (www.expasy.org) to compare the sequences of the MCM2-7 subunits of different eukaryotes as well as the bibliography available for the MCM2-7 complex purification in yeast or fruit fly to decide cloning strategy for the over-expression and purification of the hMCM2-7 complex. For that, the Mcm2 subunit was tagged with HA-tag and Strep-tag whereas the Mcm4 was tagged with FLAG-tag, 8x His-tag and TEV cleavage site.

The fusion of the pSpL2+4, pUCDM3+5 and pFL6+7 plasmids was performed following the MultiBac Expression System protocol (Fitzgerald Daniel J. et al., 2006) (see Materials and Methods, pag. 66). Sf21 cells were transfected using the bacmid generated carrying all six hMCM2-7 subunits (**Figure 11**).

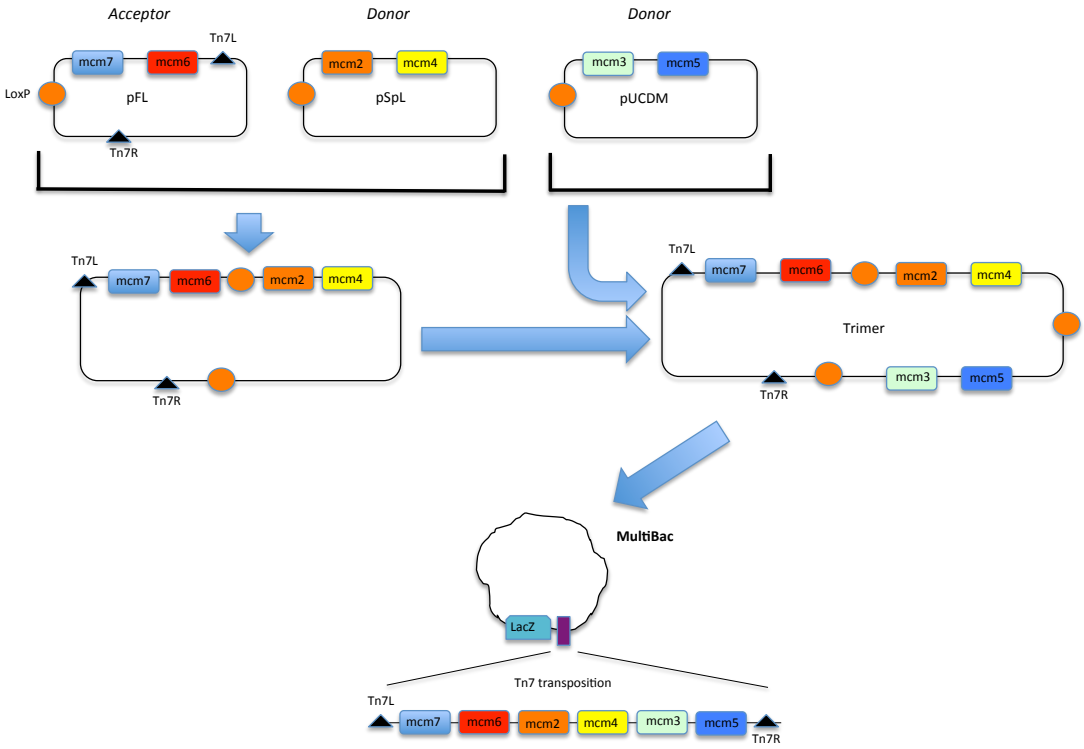



**Figure 11. MultiBac expression system and Bacmid generation**. Plasmids containing the genes that encode the Mcm6-Mcm7-Mcm2-Mcm4 were assembled firstly by LoxP site. Then the plasmid containing Mcm3-Mcm5 were fused to the pFL3-5+pSpL2+4 plasmid forming a single multigene expression construct. The trimer was fused into the MultiBac baculovirus by Tn7 transposition (Tn7L, Tn7R). This bacmid was used for infecting Sf21 insect cells and expressing the hMCM2-7 complex.

## 4.2. Over-expression and purification of hMCM2-7 complex

After transfection of Sf21 insect cells, the resulting virus 0 (V0) was used for infecting 25 ml of Sf21 at $0.5x10^6$ cells/ml to obtain Virus 1 (V1). hMCM2-7 protein expression was verified by Western blots (data not shown). V1 was titrated and used for the second generation of virus (V2). After infection, the cells were collected at DPA+48 (48 hours after Day of Proliferation Arrest), pelleted and sonicated in presence of lysis buffer (see Material and Methods). The protein over-expression was checked in 12% SDS-Page stained with Simply Blue (**Figure 12 A-B**)

For complex purification (see Materials and Methods, pag. 71), 5 liters of Sf21 insect cells were infected. After sonication of the cells, the non-soluble fraction was removed by ultra-centrifugation and the soluble fraction was incubated with Ni-NTA resin (Quiagen). The protein was eluted with an imidazole gradient. The eluted fractions were further purified using an ionic exchange HiTrap SP FF 5 ml column (GE Healthcare) and the flow-through was loaded later

into a HiTrap Heparin HP 5 ml column (GE Healthcare). Finally the complex eluted from the Heparin column was concentrated using a Vivaspin20 filtration unit (Sartorius Stedim Biotech) and loaded into a gel filtration column Superdex200 16/60 (GE Healthcare) (**Figure 12 B**).

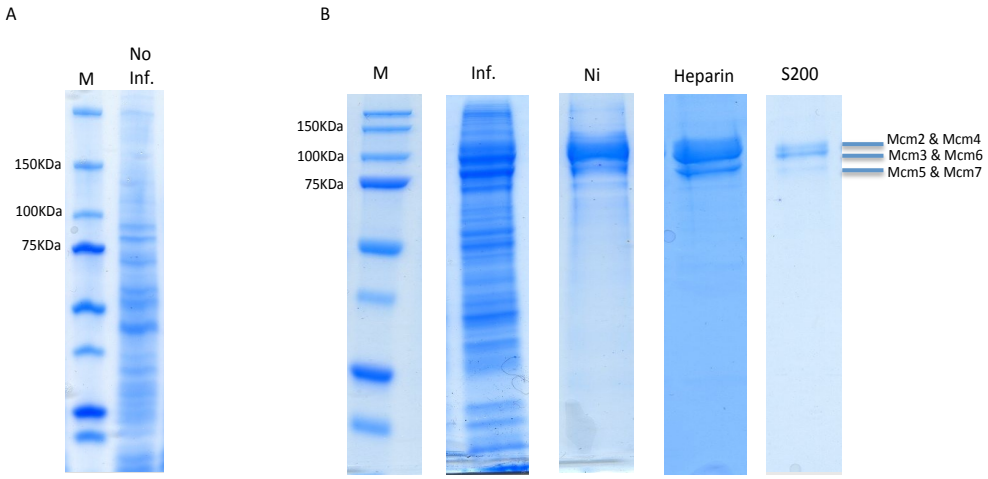



**Figure 12. Expression and purification of the hMCM2-7 complex**. (**A**) Cell lysate corresponding to non-infected Sf21 cells loaded on 12% SDS-Page and stained with Simply Blue. M: Protein marker, No Inf.: Non-infected Sf21 cells. (**B**) The different purification steps were analyzed in 12% SDS-Page. The first column (M) corresponds to the protein marker with the molecular weights indicated on the left. The Inf column corresponds to the infected SF21 overexpressing the hMCM2-7 complex. The 20 ul elution fraction analysis of the different purification steps is shown in the columns Ni (affinity column), Heparin (anion exchange chromatography) and S200 (size exclusion chromatography, Superdex 200 16/60). The identities of the bands are indicated on he right side. Note that although two bands at the non-infected cells, between the 100 kDa and 75 kDa marker, may overlap with the overexpression of the subunits Mcm3, Mcm5, Mcm6 and Mcm7, the band that appear over the 100 kDa marker, corresponding to Mcm2 and Mcm4, suggests the overexpression of the complex.

The hexamer hMCM2-7 eluted from the gel filtration column as a single peak at XXX kDa. A final peak at low-molecular weight eluted from the column and corresponds to the ATPγS present in the sample buffer (**Figure 13 A**). We did not observe a peak that could correspond to the single subunits as long as 10% glycerol was present in the buffers during the purification. The presence of the six different subunits was confirmed by western blots using specific antibodies for each subunit (**Figure 13 C**).

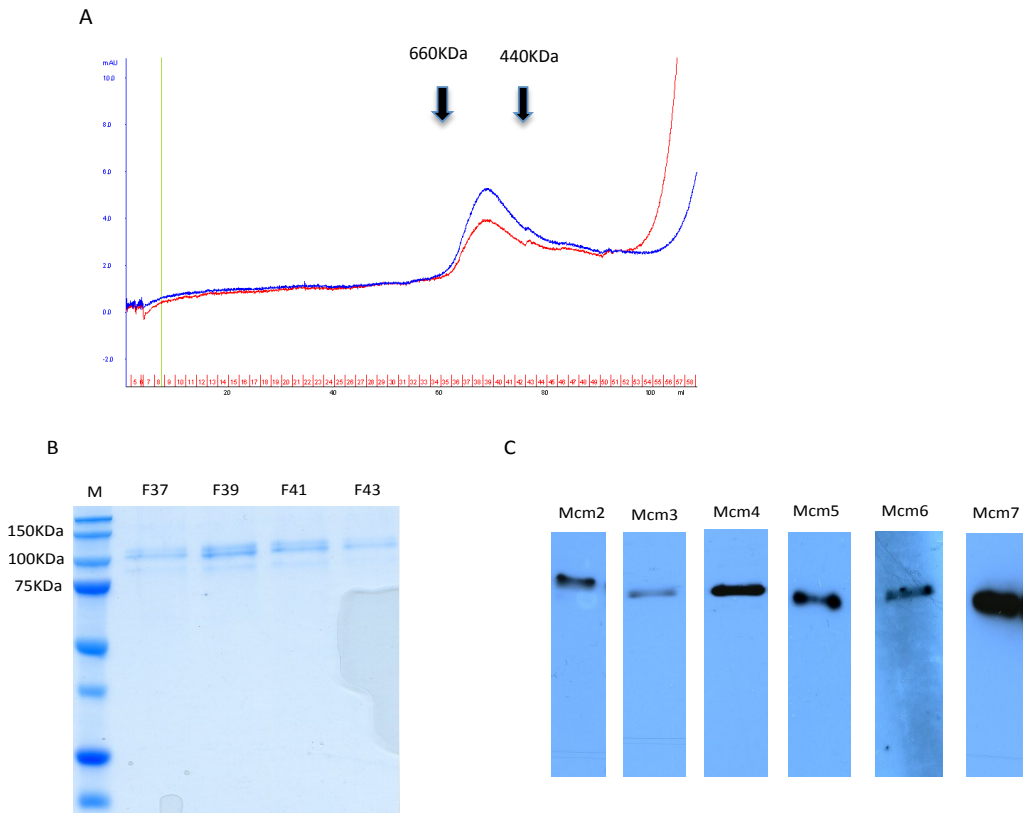



**Figure 13. Analysis of the hMCM2-7 complex**. (**A**) Chromatogram corresponding to the elution from the gel filtration column Superdex 200 16/60. A unique peak elutes from the column at 550 kDa. The high absorbance at the end of the chromatogram corresponds to the ATPγS. The molecular weights above the chromatogram correspond to Thyroglobulin (660 kDa) and Ferritin (440 kDa). (**B**) Analysis in 12% SDS-Page, stained with Simply blue, of the fractions collected during the gel filtration. The column M corresponds to the protein marker used and the molecular weights are indicated on the left. The rest of the wells correspond to the central fractions of the peak, 20 ul of protein were loaded. (**C**) Analysis by western blots of the different Mcm subunits using the fraction F39. All Mcm subunits were present without degradation. The identification for each band is indicated.

The molecular weights of the hMCM2-7 subunits are: Mcm2: 102 kDa (+2 kDa tag), Mcm3: 90 kDa, Mcm4: 96 kDa (+ 4 kDa tag), Mcm5: 82 kDa, Mcm6: 93 kDa, Mcm7: 81 kDa. When the hMCM2-7 subunits were analyzed in a 12% SDS-Page or AnyKD SDS-Page (Bio-Rad), they tended to appear clustered in three bands due to their similar molecular weights. The upper band corresponded with Mcm2 and Mcm4 subunits, the middle band with Mcm3 and Mcm6 and the lower band with Mcm5 and Mcm7. In order to separate these bands, different types of gels have been tested for the identification of the six hMCM2-7 subunits (**Figure 14 A-E**).

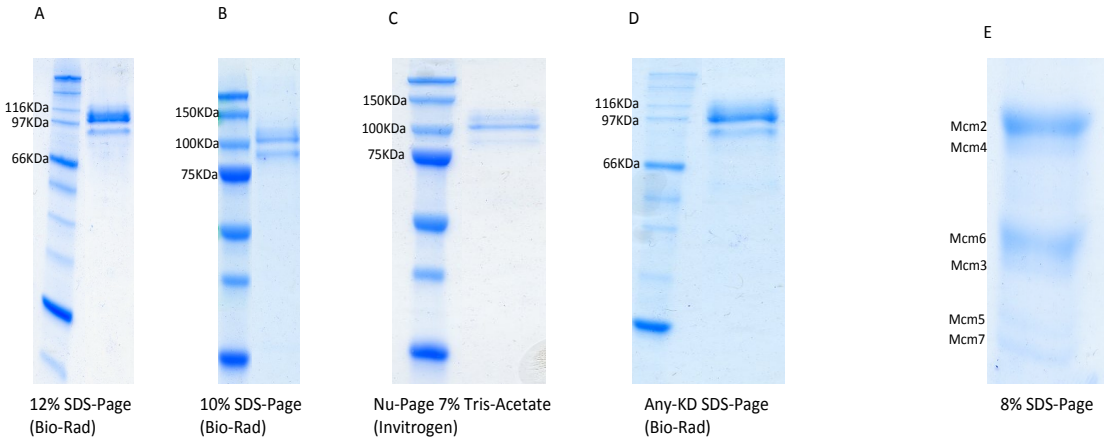



**Figure 14. Analysis of the hMCM2-7 complex using different gels in denaturing conditions**. 20 ul of hMCM2-7 complex at 0.1 mg/ml were loaded in each type of gel. (**A-D**) The molecular weights and the type of gel are indicated for each gel. (**E**) 8% SDS-Page, where the six bands can be separated. The band corresponding to each subunit is indicated.

## 4.3. Biochemical characterization of the hMCM2-7 complex

### 4.3.1. ATPase activity

All the subunits of the hMCM2-7 complex belong to the AAA+ ATPase protein family. This protein machine uses the hydrolysis of ATP to unwind dsDNA (Tuteja and Tuteja, 2004). To analyze this activity, we used a spectrophotometric technique based on the regeneration of hydrolyzed ATP coupled to the oxidation of NADH. Pyruvate-Kinase (PK) uses this ADP molecule to convert phosphoenol-pyruvate (PEP) in pyruvate, finally pyruvate is used as substrate for the L-lactate-dehydrogenase (LDH) to convert it to lactate as final product of the coupled reaction by the oxidation of one molecule of NADH (Panuska & Goldthwait, 1980) (**Figure 15**). The amount of hydrolyzed ATP is measured in an indirect manner by the NADH absorbance at 340 nm using a spectrophotometer. The absorbance of NADH decreases according to its oxidation to NAD+. The reaction is equimolar for each ATP molecule one NADH is oxidized.

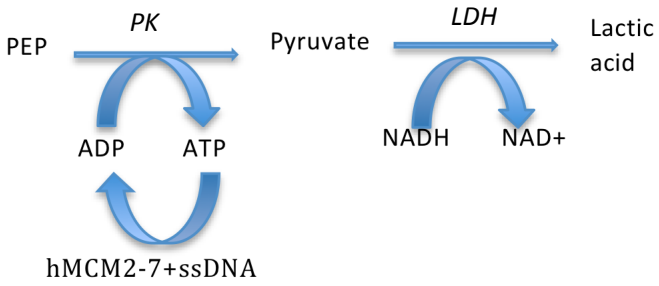



**Figure 15. Schematic representation of the couple reaction of lactate generation by PEP and ATP hydrolysis.** The absorption of NADH at 340 nm in a spectrophotometer is used for the analysis of ATPase activity.

Helicases are able to hydrolyze the γ phosphate at the ATP molecule and in many of them, this activity is enhanced when ssDNA is present, as observed for the BcMCM complex that has a viral origin (Sanchez-Berrondo J. et al., 2012). The eukaryotic MCM2-7 complex lacks this dependency (Davey et al., 2003). Taking into account that ATPase activity of BcMCM is stimulated in presence of ssDNA but not with dsDNA (Sanchez-Berrondo J. et al., 2012), the ATPase activity of the hMCM2-7 was carried on using overhangDNA (see Materials and Methods, pag. 74). As negative control we repeated the assay using the non-hydrolyzable ATPγS (**Figure 16**)

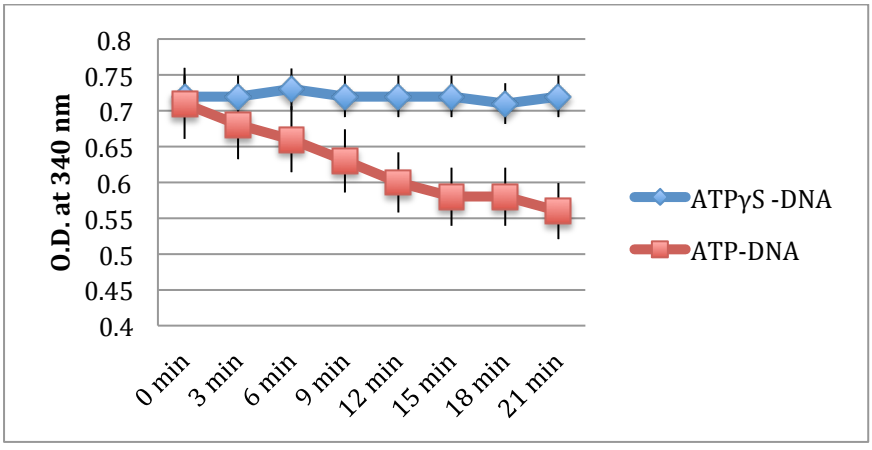



**Figure 16. Quantification of ATPase assay**. The hMCM2-7 complex was incubated with an overhang DNA in presence of ATP or ATPγS. The O.D. was measured every three minutes.

### 4.3.2. Helicase assay

The MCM complexes are replicative helicases that, by using the energy from the ATP hydrolysis, are able to unwind the dsDNA. The helicase activity of the hMCM2-7 was analyzed by electrophoresis in non-denaturing acrylamide gels, testing ssDNA displacement

from the substrate dsDNA. The helicase activity of the MCM complexes is anion-dependent, and it has been reported that the presence of glutamate enhances such function *in vitro* (Kaplan et al., 2003; Bochman & Schwacha, 2008) (see Materials and Methods, pag. 75).

It is known that the hMCM2-7 complex has 3'-5' helicase activity. For the helicase assay, the 3' overhang oligonucleotide was incubated with different hMCM2-7 concentrations and in presence of 150 mM glutamate. Importantly, the reaction was carried on in presence of magnesium acetate instead of magnesium chloride due to its helicase inhibition (**Figure 17**).

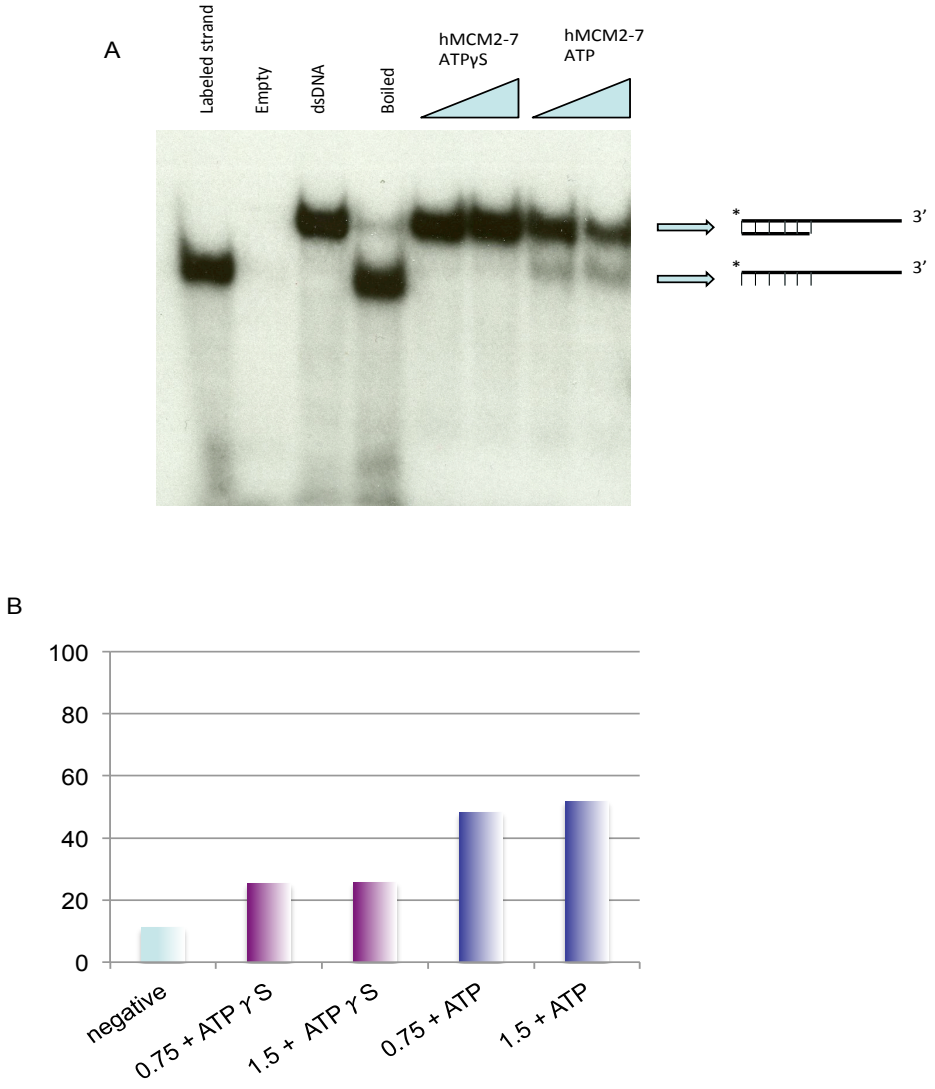



**Figure 17. Helicase assay of the hMCM2-7**. (**A**)The hMCM2-7 was incubated, at 7 nM or 15 nM, with 0.5 nM 3' overhang DNA structure. As negative control, the non-hydrolysable ATPγS homolog was used. The BcMCM helicase was used as positive control. The resulting DNA structures before/after helicase assay are indicated on the right side. The sample in each lane is indicated at the upper part. (**B**) Quantification of the helicase assay using 0.75nM or 1.5nM of hMCM2-7 complex in presence of 5mM ATPγS or ATP.

As expected, the hMCM2-7 is not able to displace the dsDNA in presence of the non-hydrolysable ATPγS,. The hMCM2-7 complex alone was able to unwind around the 30% of dsDNA *in vitro*. It is known that hMCM2-7 needs other cofactors as Cdc45 and GINS complex to display full activity.

### 4.3.3. DNA binding assay

The helicases usually associate with higher affinity to ssDNA rather than dsDNA (Sanchez-Berrondo J. et al., 2012). For the purification of the hMCM2-7/DNA complex, we have used magnetic beads covered with streptavidin and an overhang DNA labeled with biotin (see Material and Methods, pag. 72). Incubating the biotinilated DNA with the streptavidin coated magnetic beads (Dynabeads M-280 Streptavidin, Invitrogen), the DNA interacts with the beads in a specific and stable manner. Thus, the labeled DNA was retained on the beads together with the hMCM2-7 complex bound to the ssDNA part. The hMCM2-7/overhangDNA complex was released from the beads upon cutting the DNA with EcoR I restriction enzyme.

Following this procedure we avoided mixtures of complexes with and without DNA. For that, we generated 60bp overhang DNA that was coupled to biotin molecule at 5' and includes an EcoR I restriction site and a 49 poly-T 3' ssDNA fragment. The DNA was incubated with the beads at room temperature for one hour. Then, hMCM2-7 complex was attached to the beads+DNA in presence of 5 mM ATPγS and washed several times to eliminate material unbound to the DNA. Finally, the complexes bound to DNA were released from the beads by incubating over-night with EcoR I enzyme (**Figure 18**). The same reaction was carried out without DNA as a negative control, to show that no specific interaction occurs between the protein complex and the beads. The samples obtained were visualized in SDS-PAGE and stained with Simply blue.

This experiment allowed us to obtain a homogeneous hMCM2-7/overhangDNA sample used to generate the 3D structure of the complex by negative-stain electron microscopy.

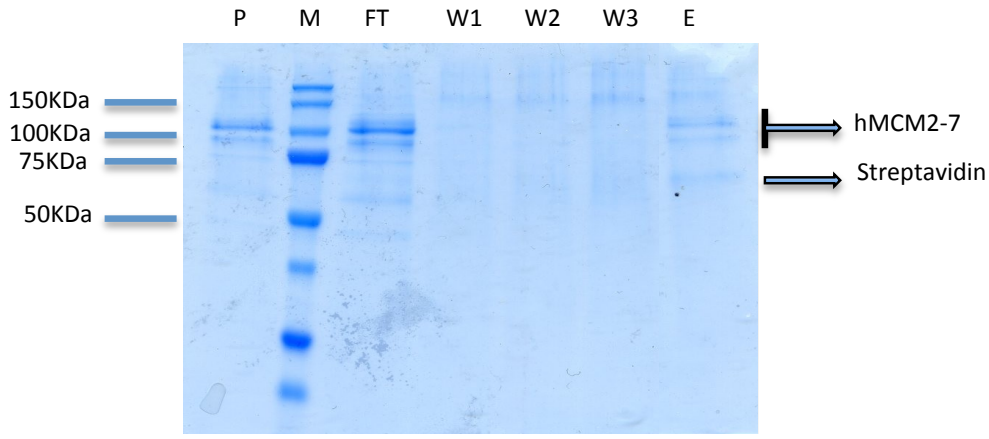



**Figure 18. Purification of the hMCM2-7 complex together with overhangDNA**. The well P, is the hMCM2-7 pure; M: marker with the molecular weights indicated on the left of the gel; FT: Part of the protein not bound to the beads-DNA; W1-W3 correspond to the washes; E: elution fraction. The eluted proteins are indicated on the right side of the gel.

## 4.4. Structural characterization of the hMCM2-7 complex

### 4.4.1. 2D analysis of hMCM2-7 APO and hMCM2-7 bound to ADP

Initially, the hMCM2-7 complex was firstly analyzed in its apo state in order to visualize the putative conformational changes upon nucleotide binding. Thus, the purification of the hMCM2-7 complex was performed using buffers without the addition of any nucleotide in the final gel filtration column. 3 ul of freshly purified complex were applied onto a pre-ionized carbon grid negatively stained and observed in transmission electron microscope. The EM images revealed a relatively homogeneous distribution of particles. We collected 21434 particles and use them to generate reference-free 2D averages to increase the signal to noise ratio. The 2D averages obtained are similar to the top/bottom views of other eukaryotic MCM2-7 complexes with a clear central channel whereas the side views presented the two roughly parallel layers that are similar to averages observed for other MCM2-7 complexes (Sanchez-Berrondo et al., 2012; Costa et al., 2011). However, a closer look to the top/bottom views showed high conformational heterogeneity in the sample. The side views had different lengths and at the top/bottom views, the individual subunits were hardly aligned. Therefore, the open-ring conformation of the MCM2-7 complex seems very flexible in hMCM2-7 in apo state. This fact disallowed us to generate a 3D reconstruction of the hMCM2-7 complex in absence of nucleotide (**Figure 19 A**) (see Materials and Methods, pag. 76-77).

A fraction of purified hMCM2-7 complex in the absence of nucleotide was then incubated for 2 hours in presence of 10 mM ADP and subsequently negatively stained. 10370 articles were manually collected and used to generate reference-free 2D classes. The 2D averages of the hMCM2-7 complex presented the same views as the apo state and although the top/bottom views suggested a more constricted structure, we could not generate a 3D structure due to the still high conformational heterogeneity (**Figure 19 B**).

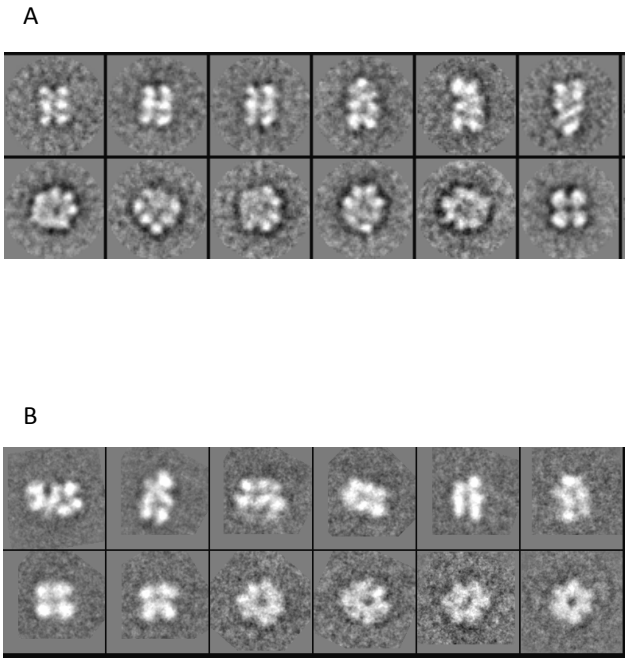



**Figure 19. Reference-free 2D classification of the hMCM2-7 in apo state and 10mM ADP-bound state**. (**A**) Reference-free 2D classes of the apo hMCM2-7. The top/bottom views present some well-defined subunits but also some blurry areas due to the flexibility of the complex. (**B**) Reference-free 2D classification of the hMCM2-7 with 10 mM ADP. The top/bottom views are better defined than in the apo state but still too heterogeneous to generate a unique 3D model.

### 4.4.2. 3D reconstruction of the negatively-stained hMCM2-7 bound to ATPγS

To obtain the 3D structure of the hMCM2-7 complex, 5 mM ATPγS was added to the complex before loading into the gel filtration column. The fractions coming from the Superdex 200 column were also incubated with 5 mM ATPγS and loaded onto a pre-ionized grids covered with thin carbon layer. These grids were stained with 2% uranyl-formate. Several hundreds of low-dose micrographs were processed to get an initial data set of 20394 particles. The references free 2D classes averages and 3D reconstruction were done by using the best 10494 particles and processed with EMAN1.9 and Xmipp packages (**Figure 20 A-B**) (see Materials and Methods, pag. 76-77).

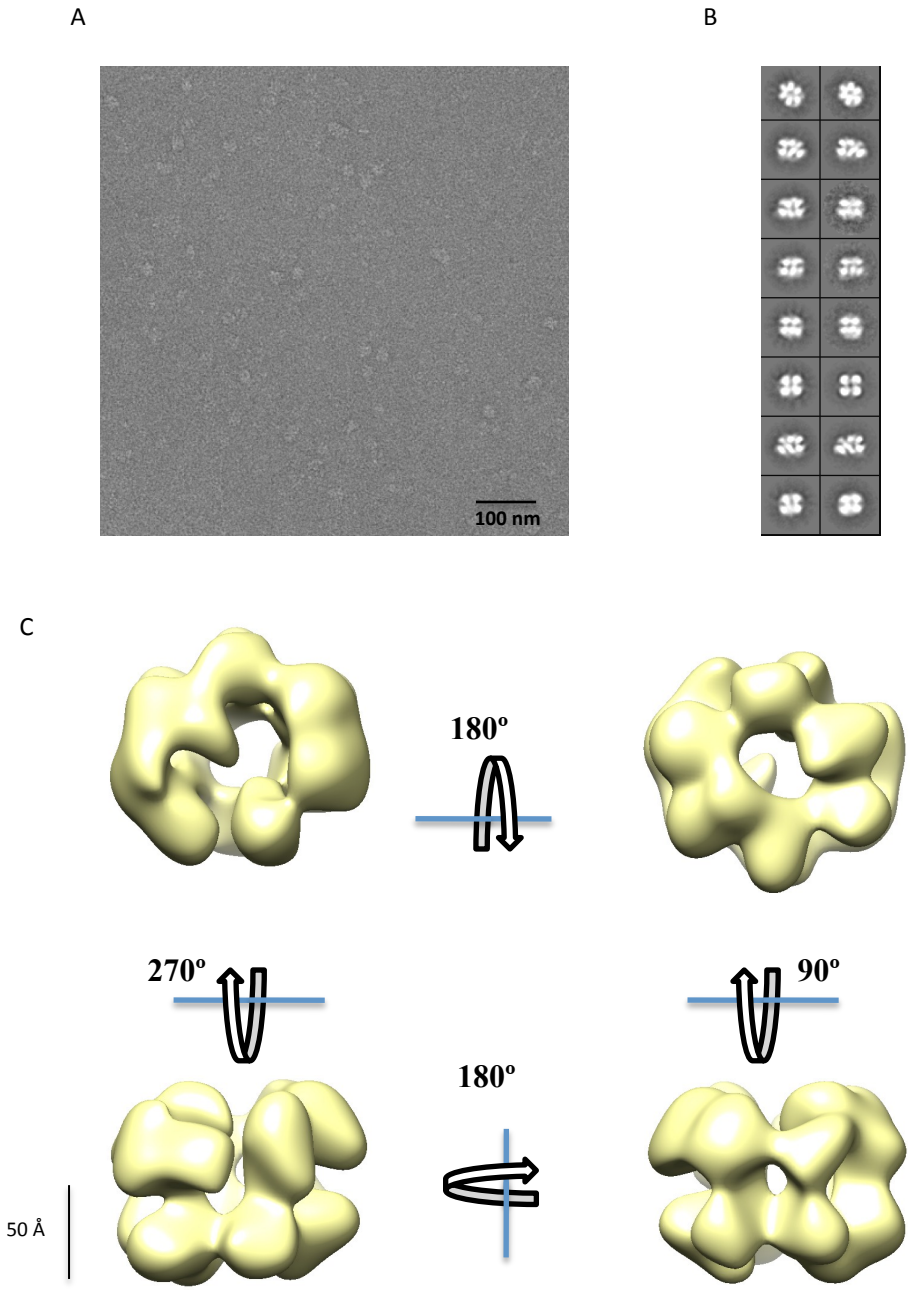



**Figure 20. 3D reconstruction of the human MCM2-7complex with ATPγS** . (**A**) Sample electron micrograph of negatively stained hMCM2-7 complex with 5 mM ATPγS. (**B**) Comparison of the hMCM2-7 complex reference free 2D classes averages (left column) matching its model projections (right column). (**C**) Different views of the final 3D reconstruction of the hMCM2-7 complex bound to 5 mM ATPγS at 24 Å resolution.

The 3D structure of the hMCM2-7 complex displays a classical hexameric shape and keeps some similarities with the *Drosophila* MCM2-7 complex (Costa et al., 2011) (**Figure 20 C**). The human complex revealed an open-ring structure at the N-terminal side whereas the C-terminal part is closed. The structure was solved at 24 Å resolution, which allow us to

distinguish the boundaries between the different subunits. The size of the hMCM2-7 complex bound to ATPγS is 100 Å high, 130 Å large and 130 Å deep. The gap between the Mcm2 and Mcm5 has distance of approximate 15 Å whereas the central channel displays a 30 Å diameter. The six smaller side-channels that exist between the hMCM2-7 subunits are also well visible in our model. The lateral views of this complex are divided in two roughly parallel layers of high-density corresponding to the C-terminal and N-terminal domains of the Mcm subunits, whereas the low-density central part is formed by a large loop that links these two domains.

### 4.4.3. 3D reconstruction of the negatively-stained hMCM2-7 complex bound to ATPγS and overhangDNA

The sample obtained from the purification with the magnetic beads was applied on pre-ionized grids covered with carbon layer. After staining them negatively with uranyl formate, the complex was visualized under the electron microscope. To perform the 3D reconstitution of the complex with DNA, we collect 18268 particles for hMCM2-7–ATPγS – overhangDNA and 11195 of them were used for the processing (see Materials and Methods, pag. 76-77). The reference free 2D averages corresponding to lateral views present similar characteristics than the complex with ATPγS alone, however at the top/bottom view the six subunits look more defined and the complex seems more compact, compared to the complex without DNA. Due to the physical limitations of the negative staining technique, the overhangDNA was not visible in the structure (**Figure 21 A-B**).

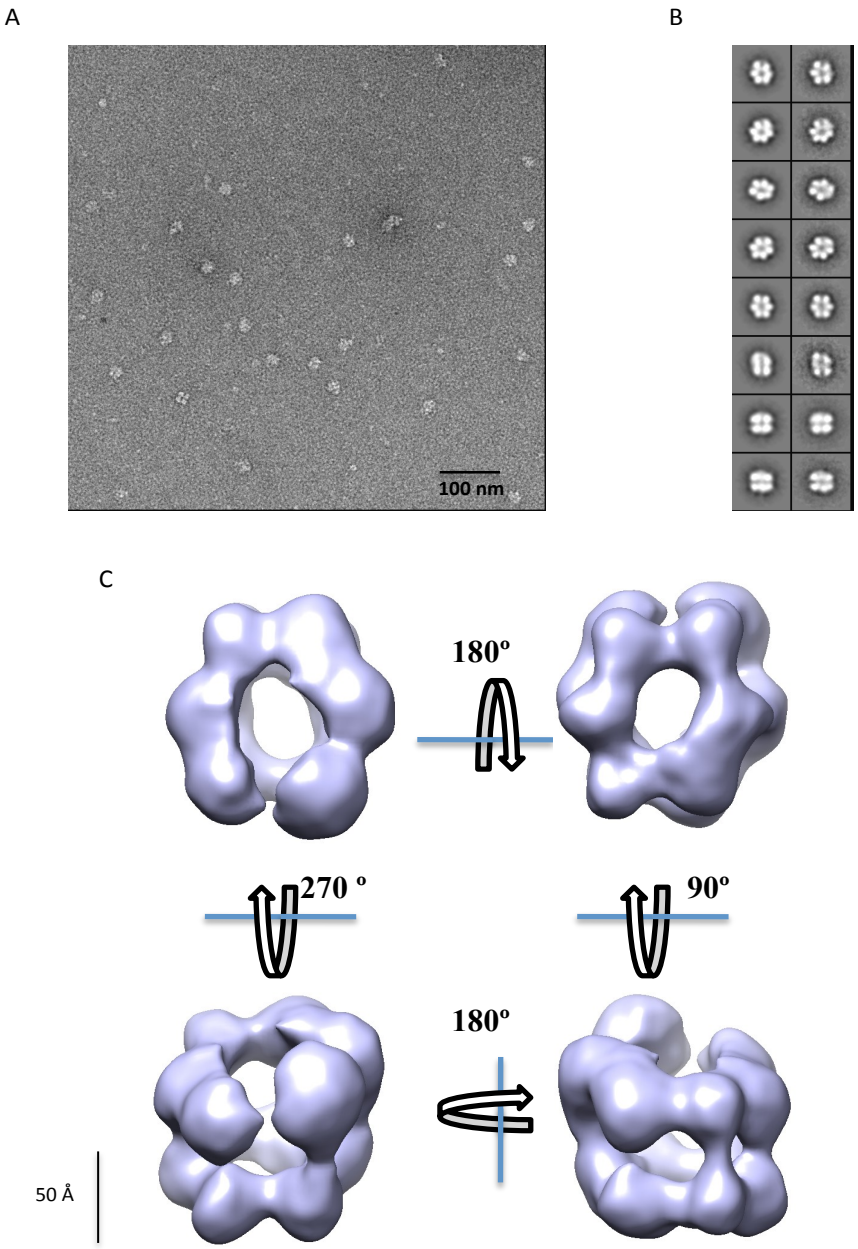



**Figure 21. Structure of the human complex MCM2-7 with overhangDNA**. . (**A**) Sample electron micrograph of negatively stained hMCM2-7 complex bound to overhangDNA. (**B**) Comparison of the hMCM2-7 complex reference free 2D classes averages (left column) matching its model projections (right column). (**C**) Different views of the volume of the hMCM2-7 complex with overhangDNA at 28 Å resolution.

The dimensions of the hMCM2-7 complex bound to the overhangDNA and with six-fold symmetry applied, are 110 Å high, 125 Å large and 130 Å deep, revealing a more compact structure than the hMCM2-7 complex without DNA (**Figure 21 C**). The central channel has become narrower when comparing it with the MCM2-7 without DNA. The complex still presents the gap between the Mcm2 and Mcm5 although is significantly narrower than

compared with the complex in absence of DNA. The size of the complex has changed as well, being now larger and narrower than the complex without DNA.

### 4.4.4. Cryo-electron microscopy. 2D analysis of hMCM2-7 bound to overhangDNA

With the aim to improve the resolution of the hMCM2-7 complex bound to the overhangDNA and visualize the DNA, we performed cryo-electron microscopy. For that the sample obtained by EcoR I elution from the beads was firstly concentrated twenty times and then diluted 1:8 in order to decrease the glycerol down to 2%. The sample was loaded on cryo-EM grids, covered with thin carbon film and incubated for 4 minutes before flash freezing in ethane-near liquid nitrogen temperature (see Materials and Methods, pag. 78). 2000 micrographs were collected. From an initial data set of 6500 particles manually selected, we could only use 850 particles to generate the reference free 2D classes averages due to the bad quality of the dataset. Although the number of particles is very low to determine 3D structure by cryo-EM, we could generate some reference free 2D classes averages (**Figure 22 A-B**).

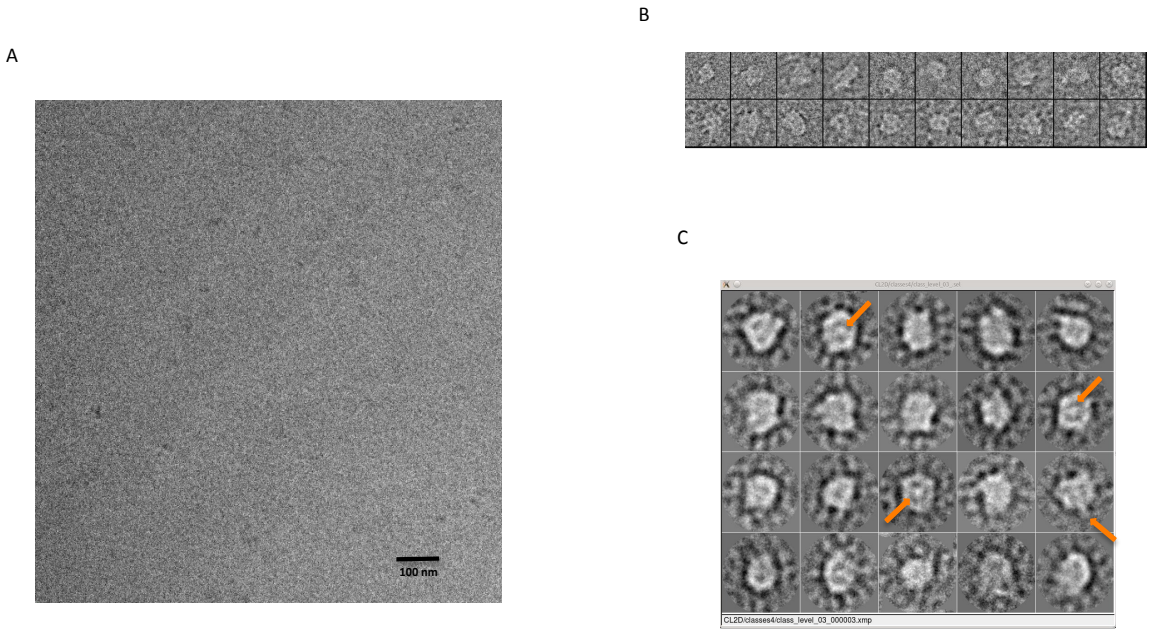



**Figure 22. Cryo-electron microscopy of the hMCM2-7 complex with overhangDNA**. (**A**) Sample cryo-electron micrograph of hMCM2-7 complex bound to overhangDNA. (**B**) Gallery of single particles data set collected from 200 images. (**C**) Reference free 2D classes averages from 850 particles using CL2D. The red arrows show an extra density in the middle of the ring as well as some density coming out from the molecule in the lateral view that could correspond to DNA.

Some of the top/bottom views of the hMCM2-7 complex present a density at the position of the central channel that might correspond to the DNA. Also some density seems to come out from the complex when looking at the lateral views indicating the possible presence of dsDNA (**Figure 22 C**)

## 4.5. CMG complex

The CMG is composed by Cdc45 and the MCM2-7 and GINS complexes. This supra-assembly constitutes the core of the eukaryotic replisome, acting as the replicative helicase *in vivo REF*. For the assembly of the CMG complex *in vivo*, the MCM2-7 must be firstly loaded onto the dsDNA, then Sld3 will carry Cdc45 to MCM2-7 and the interaction of TopBP1 with Sld3 facilitates the assembly of the GINS complex and Polymerase ε. Finally, the release of some proteins, among them Sld3 and TopBP1 will give rise to the formation of the CMG complex, which together with the polymerase ε, form the Replication Progression Complex (Gambus et al., 2006; Ilves et al., 2010).

It is known that the CMG complex only binds ssDNA (Remus et al., 2009; Ilves et al., 2010). Also the checkpoints that cell uses to pass through each phase of the cell cycle remain essential for the CMG formation. It is known that phosphorylations in different MCM2-7 subunits by Cyclin-dependent Kinases (CDK) and Dbf4-dependent Kinases (DDK) are necessary for the assembly of the CMG complex (Im et al. 2009). Cdc7, a DDK protein, phosphorylates the amino-terminal domain of the Mcm2, Mcm4 and Mcm6 increasing the stability of the binding between the complex MCM2-7 and Cdc45 protein (Zegerman et al., 2007; Labib et al., 2010). For CDK is known that phosphorylates Sld2 and Sld3, necessary for the recruitment of GINS complex to the MCM2-7 and Cdc45 (Diffley et al., 2007).

### 4.5.1. hCMG complex assembly

In a first attempt to reconstitute the CMG complex, the over-expressed hMCM2-7 was loaded onto the overhangDNA bound to streptavidin-coated beads and then human Cdc45 and hGINS complex were incubated for 4 hours to allow the assembly of the CMG components. After removing the unbound proteins fraction by washing the beads with Superdex buffer, the ssDNA was eluted together with the bound proteins upon DNA cleavage with EcoR I. The elution fraction was analyzed in a SDS-Page. The proteins were

visualized by silver staining (**Figure 23**).

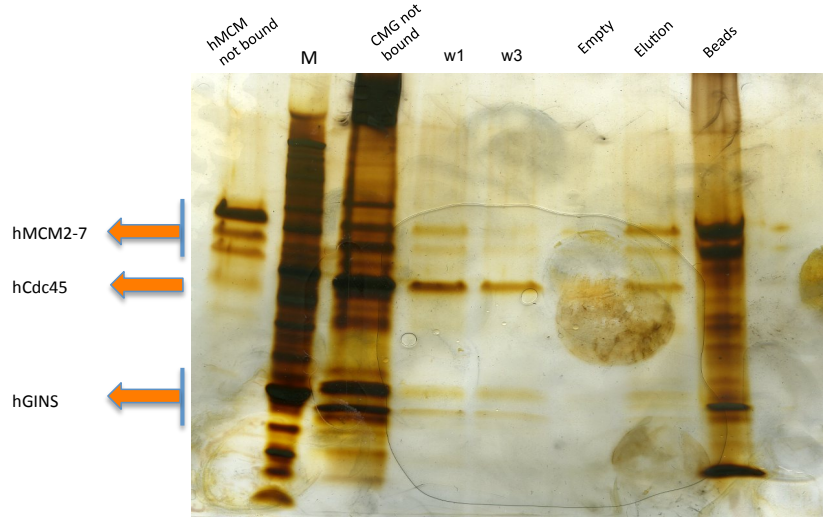



**Figure 23. Assembly of the CMG complex**. The bands corresponding to each component of the CMG are indicated on the left. M: marker; W1-W3: first and third wash fractions. CMG not bound: Corresponds with the hCdc45 and hGINS complex not bound to the DNA nor to the hMCM2-7; Beads: Fraction corresponding to the proteins bound to the magnetic beads. The beads were incubated at 94ºC for ten minutes in presence of 6x Protein Loading dye to remove any protein bound.

The incubation of the hMCM2-7 loaded onto ssDNA with hCdc45 and hGINS complex in the gel filtration buffer did not yield an assembled CMG, most likely due to the lack of a cofactor or regulatory posttranslational modifications. Most of the hGINS complex added to the assay, did not bind to the hMCM2-7. During the three washes steps, some hMCM2-7 was eliminated together with Cdc45 and the remaining hGINS. The hMCM2-7 complex and Cdc45 are the main components of the elution fraction. The weak bands corresponding to the hGINS complex in the elution fraction may arise from a part of this complex bound to the ssDNA as it has been previously reported (Boskovic et al., 2007).

The experiment was repeated, and the elution fraction coming from the beads was injected in an analytical gel filtration column Superdex 200 10/300 to separate the possible different species (**Figure 24**). Proteins were separated in a SDS-Page and visualized by stain silver staining.

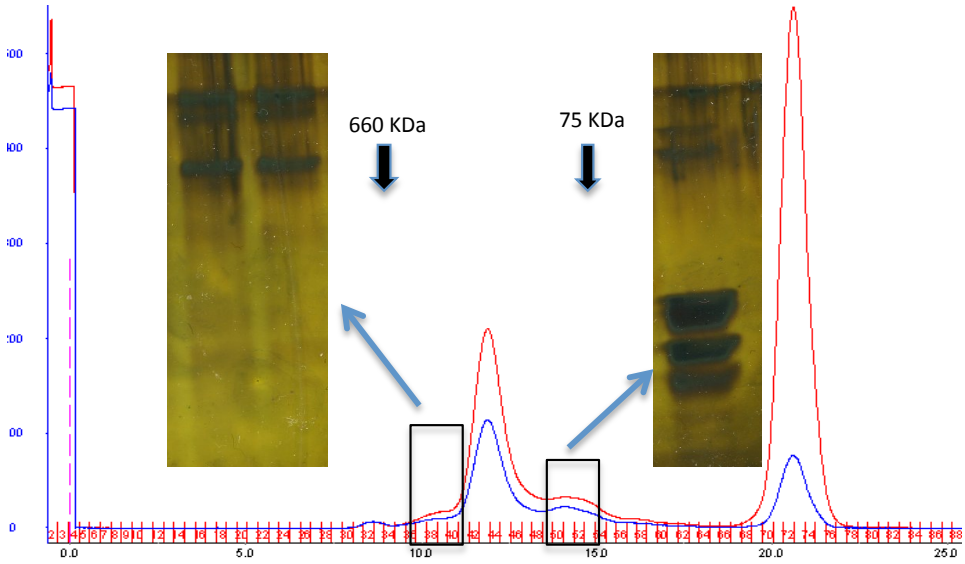



**Figure 24. Assembly of the CMG complex and purification.** The shoulder coming before the main peak of the chromatogram, seemed to contain both hMCM2-7 and Cdc45, and was corresponding with the theoretical molecular weight of a complex hMCM+Cdc45. The peak corresponding with hGINS complex was also labeled. Molecular weights for Thyroglobulin (660 kDa) and Conalbumin (75 kDa) are indicated above the chromatogram. The central peak corresponds to hMCM2-7.

These two experiments suggested us the possibility that hGINS complex is bound less tightly to the CMG complex than Cdc45 or hGINS only binds ssDNA. However, these experiments opened the possibility to form a complex between hMCM2-7 and Cdc45.

### 4.5.2. Co-expression of hMCM2-7 and hCdc45

To improve the assembly of the hMCM+Cdc45 complex, the hMCM2-7 and the human Cdc45 protein were co-expressed and purified together. Sf21 insect cells were co-infected with hMCM2-7 and hCdc45 baculoviruses. For the purification of this CMG subcomplex the purification protocol for the hMCM2-7 was followed. The purification was done in presence and absence of ATPγS to discard possible conformational changes that could inhibit the binding between hMCM2-7 and Cdc45. Although along the purification, hMCM2-7 complex and Cdc45 were co-eluting along the purification (**Figure 25 A**) when passing the sample through the gel filtration column, Cdc45 eluted at the end of the column, several fractions away from the hMCM2-7 elution peak (**Figure 25 B**). The proteins were visualized in SDS-Page.

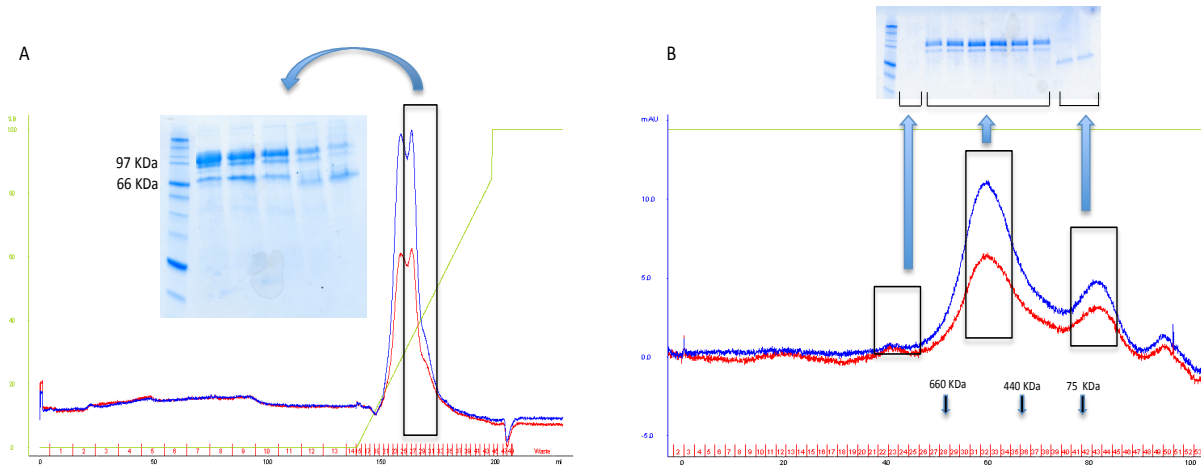



**Figure 25. Co-expression of hMCM2-7 complex and human Cdc45 protein**. (**A**) Chromatogram of the heparin column. At the indicated peak, the hMCM2-7 and Cdc45 were co-eluting. (**B**) Chromatogram of the Superdex 200 16/60. The elution peaks correspond to the fractions visualized in the SDS-Page. Cdc45 elutes separated from hMCM2-7. The molecular weights of Thyroglobulin (660 kDa), Ferritin (440 kDa) and Conalbumin (75 kDa) are indicated under the chromatogram.

We could not reconstitute the hMCM2-7+Cdc45 complex by incubating the proteins in presence of DNA and ATP. It's known by the structure of the Drosophila CMG complex that Cdc45 together with GINS encloses the gate in between Mcm2 and Mcm5 (Costa et al., 2011). Probably GINS is required to stabilize the binding between Cdc45 and hMcm2.

In this experimental condition, the lack of CDK and DDK, and other post-translational modification proteins required for CMG assembly, made *in vitro* reconstitution of that complex not possible in these conditions.

### 4.5.3. *In vitro* reconstitution of the human CMG complex

To understand the molecular basis of the CMG complex assembly, we have recreated the CMG *in vivo association* in the G1/S phase.

HeLa cells were synchronized at G1/S phase and a cell extract was prepared (see Materials and Methods, pag. 79-80). Different biotin-labeled DNA oligonucleotides were designed to study the loading of the CMG complex: a 3' overhang DNA, a Bubble DNA forming a structure similar to two opposite replications forks and a Fork DNA mimicking a single replication fork (see Material and Methods, pag. 80) (**Table 10**). The hMCM2-7 was loaded onto DNA, then human Cdc45 protein and hGINS complex were added in presence of G1/S synchronized HeLa cell extract. Proteins not bound to the hMCM2-7 or free DNA were

eliminated by three washes with Superdex buffer. The proteins were eluted from the beads by digestion of DNA by EcoR I restriction enzyme (**Figure 26**).

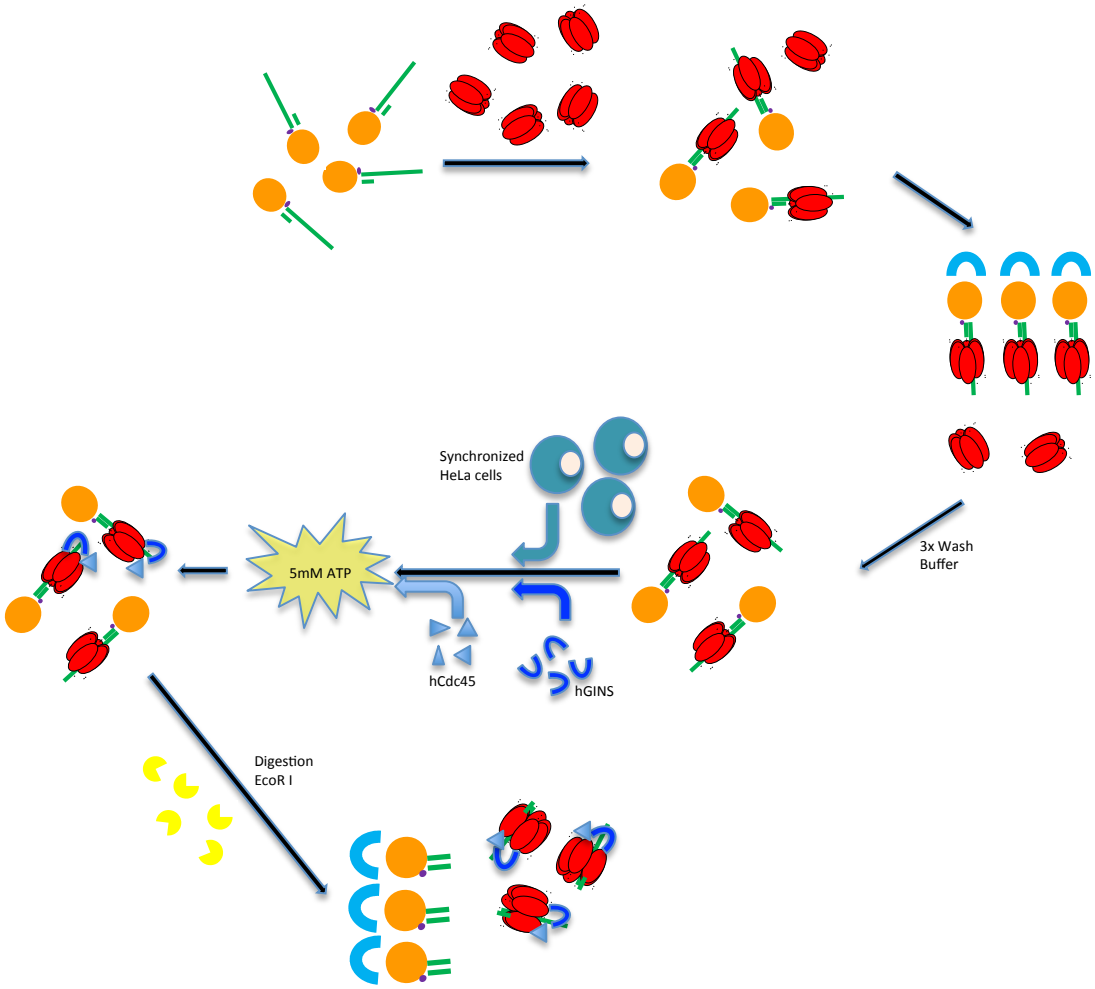



**Figure 26. Schema of the CMG reconstitution**. After loading of the hMCM2-7 onto the ssDNA, hGINS and hCdc45 were added in presence of synchronized HeLa cell extract and ATP. Unbound proteins were removed by bead washing and the elution was performed using EcoR I.

The HeLa cells extract synchronized at mitotic phase was used as negative control. The effect of the presence of ATP or a non-hydrolysable analogue, such as ATPγS was also analyzed at the CMG reconstitution assay (**Table 10**).

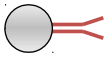
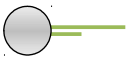
| Type of DNA | G1/S cell extract | Mitotic cell extract |
|---|---|---|
|  | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS |
|  | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS |
|  | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS | Cell extract only<br>hMCM2-7 + ATP<br>hMCM2-7 + ATPγS<br>hMCM2-7 + Cdc45 + GINS + ATP<br>hMCM2-7 + Cdc45 + GINS + ATPγS |

**Table 10. Resume of the components used for CMG reconstitution assay**.

The protein analysis was done by the Mass Spectrometry Unit at the CNIO (Madrid, Spain)(see Materials and Methods, pag. 81-82). The proteins interacting with the DNA probe were subjected to label free proteome analysis. The same protein amount was injected in duplicates to increase the confidence of the values. Raw files were searched against a human forward-reverse concatenated database (UniProtKB/Swiss-Prot) using Proteome Discoverer software and only peptides filtered at 1% FDR were considered.

After analyzing all the experiments, 2870 proteins were detected in this assay. After summing all the PSMs (Peptide Spectrum Mass) for each of the proteins, the ones with less than 80 PSMs were eliminated from the list, giving rise a total of 976 proteins (see **Figure 27 A-B**).

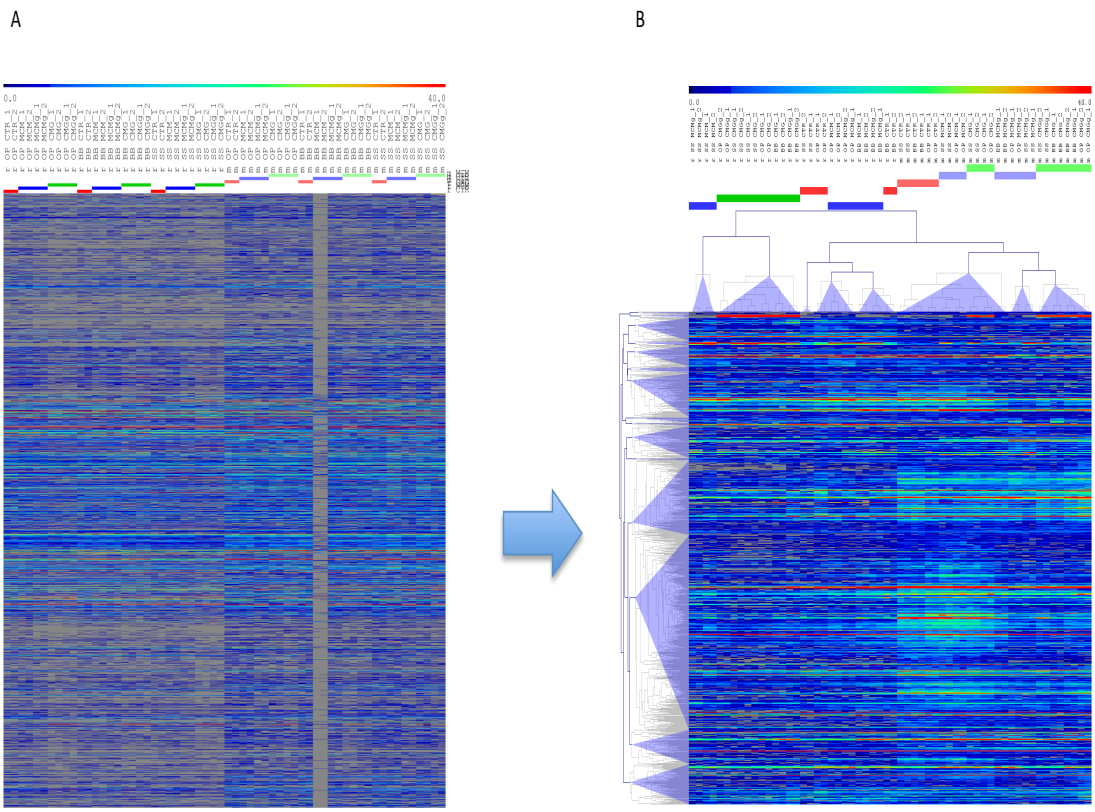



**Figure 27. Proteins detected at the CMG reconstitution assay**. Overview of proteins detected in the assay. The values of PSMs are represented in the color bar from blue (0 PSMs) to red (over 40 PSMs). (**A**) The 2870 proteins detected in the assay, the left half of the panel represents the assay done with G1/S cell extract, whether the right half was done with Mitotic cell extract. (**B**) Proteins with an overall PSMs over 80, the left half of the panel represents the assay done with G1/S cell extract, the right half was done with Mitotic cell extract. Proteins were represented with the Multiple Experiment Viewer 4.9.

Among the 976 proteins with a number of PSMs over 80, all proteins that form the CMG complex were present. The values of PSMs detected for the hMCM2-7 subunits, Cdc45 and GINS subunits were analyzed to study whether the presence of G1/S cell extract would have an effect on the CMG reconstitution compared to the presence of the mitotic cell extract (**Figure 28**).

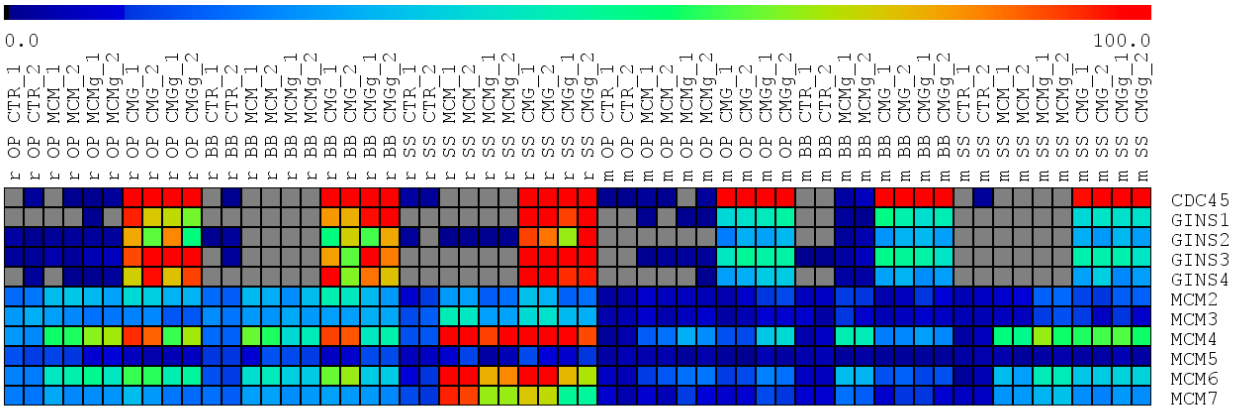



**Figure 28. Representation of the levels of PSMs detected for the CMG proteins**. The signal saturates when PSMs are over 100. R G1/S cell extract; M Mitotic cell extract; OP Fork DNA; BB Bubble DNA; SS 3' overhang DNA; CTR Control. DNA incubated with the corresponding cell extract; MCM hMCM2-7 incubated with the oligonucleotide in presence of 2mM ATP; MCMg hMCM2-7 incubated with DNA in presence of 2mM ATPγS; CMG hMCM2-7 + Cdc45 + GINS incubated with DNA in presence of 2mM ATP; CMGg CMG hMCM2-7 + Cdc45 + GINS incubated with DNA in presence of 2mM ATPγS. The chart was generated with Multiple Experiment Viewer 4.9.

The G1/S cell extract had a strong effect on the CMG reconstitution. hMCM2-7 complex was much better loaded onto the DNA in presence of G1/S cell extract (**Figure 28**) compared to the mitotic cell extract. Also GINS complex and Cdc45 protein were loaded in higher levels upon incubation with G1/S cell extract, indicating that proteins present in the replication cell extract enhances the binding of the CMG proteins (Fig. 28). Although the fork DNA is the structure more similar to a replicative fork, the CMG complex assembles better *in vitro* when long ssDNA is present (Fig 28). Using the mitotic cell extract, the levels of the hMCM2-7 and GINS loaded onto any of the oligonucleotides was much lower than the levels obtained when using the G1/S cell extract. Only Cdc45 remained in high levels, probably due to its interaction with ssDNA.

In addition, 24 DNA-damage related proteins were detected together with the reconstituted CMG complex. We decided to explore whether any of these DNA-damage related proteins was associated with the CMG complex or they may be pulled-down due to the DNA probe used in the assay (**Figure 29 A**).

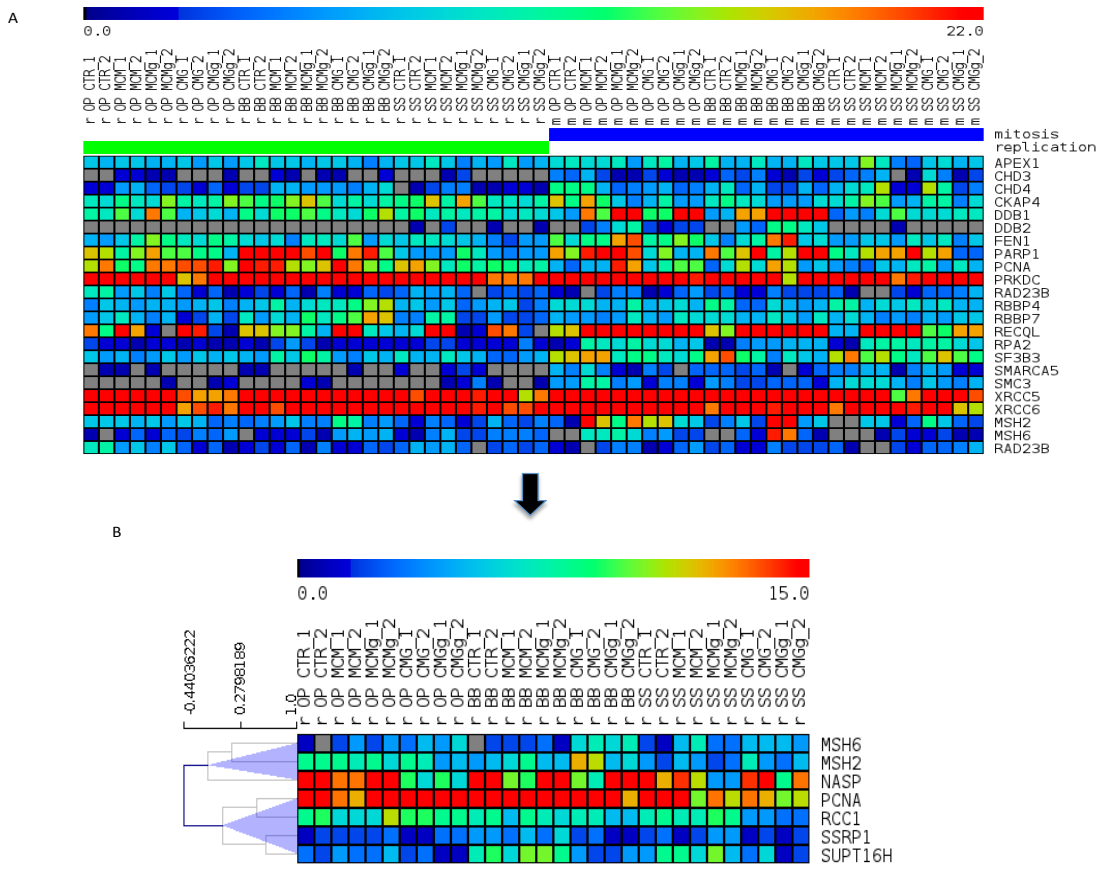



**Figure 29. DNA-damage related proteins.** (**A**) 24 DNA-damage proteins were detected in our assay, using both G1/S and Mitotic cell extract. (**B**) Pearson correlation for some of the DNA-damage related proteins at G1/S phase.

Proteins associated with the replisome are more abundant in presence of G1/S cell extract. PCNA accurately confirms replication timing.

Among the DNA-damage related proteins, the Mut-L complex formed by Msh2 and Msh6 proteins, was present in the replication assay. Its function is to eliminate mispaired bases resulting from replication errors by subsequent recruitment of the Mlh1-Pms1 complex. In our experiments, Msh2 was more abundant when reconstituting the CMG complex on the Bubble DNA and in presence of ATP. A direct interaction between PCNA and the MMR proteins Msh6 and Msh3 (Mut-S complex) was shown (Clark et al., 2000). The PSMs obtained from Msh2/6 were analyzed together with other proteins as PCNA by the Pearson correlation Factor. Its negative correlation (-0.44) leaded us to think about the possibility that the presence of the Mut-L complex would inhibit the presence of PCNA (**Figure 29 B**). The CMG complex assembled on Bubble DNA with ATP and in presence of Msh2/6 and PCNA was analyzed by Western blots using specific antibodies for each protein, indicating that Mut-L complex would co-localize together with PCNA (data not shown).

Most of the proteins that were detected in this assay have been described as replisome-associated factors analyzing the human replisome vicinity at nascent DNA molecules (Lopez-Contreras et al., 2013).

### 4.5.4. Re-purification of human CMG complex

In order to reduce the background of the proteins bound to DNA or weakly bound to the CMG complex, obtained in presence of G1/S cell extract, we used the histidine-tag at the hMcm4 subunit to pull-down the whole CMG-DNA complex and associated proteins. Thus the elution fraction after the EcoRI digestion was incubated with Ni-NTA resin. The CMG complex was eluted in one step using 500 mM imidazole after three washes at 30 mM imidazole. The same experiment was carried out omitting the replicative cell extract in the assembly step. Both experiments were analyzed by Western blotting (**Figure 30**) and by the Mass Spectrometry Unit at the CNIO (Madrid, Spain).

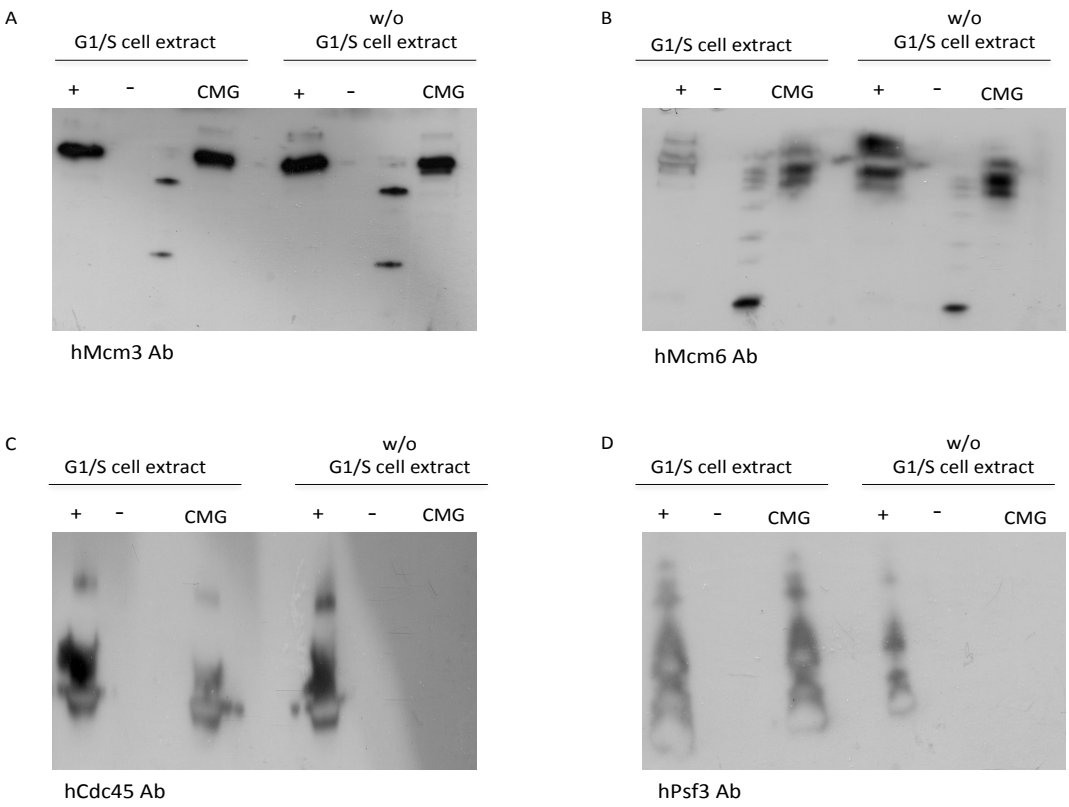



**Figure 30. Western blot analysis of purified CMG components**. The assay was done in presence/absence of G1/S cell extract. (**A-B**) hMcm3 and hMcm6 detection when the CMG complex is assembled. The loading of the hMCM2-7 on the overhangDNA does not require the cofactors present in the cell extract. (**C**) Detection of the human Cdc45 protein. Cdc45 binds hMCM2-7 only in presence of the cell extract supplemented with ATP. (**D**) Detection of the human Psf3, one of the subunits of GINS complex. The interaction of GINS with hMCM2-7 needs the cell extract to co-elute with the hMCM2-7 complex.

Therefore, the cofactors present at the G1/S cell extract are needed for the assembly the CMG complex. The loading of hMCM2-7 complex on the overhangDNA does not depend of the cell extract. However, the loading of Cdc45 and GINS on hMCM2-7 complex does require the activity of other cofactors, as CDKs and DDKs, to form the CMG complex.

### 4.5.5. Essential phosphorilations for the hCMG reconstitution

The high number of proteins detected at the CMG reconstitution assay, make difficult the detection of phosphopeptides in the sample. To decrease the "noise" of the experiment and increase the possibility to detect phosphopeptides , for proteomic analysis we used the re-purified CMG complex and analyzed the sample by the Mass Spectrometry Unit at the CNIO (Madrid, Spain) (see Materials and Methods, pag. 81-82).

A

| Position | UniProt | PhosphoSite | Subunit | Modification | Confidence |
|---|---|---|---|---|---|
| S139 | Yes | No | Mcm2 | Phospho | High |
| Y137 | No | Yes | Mcm2 | Phospho | Medium |
| T35 | No | Yes | Mcm2 | Phospho | High |
| S27/S28 | Yes | No | Mcm2 | Phospho | High |
| S672 | Yes | No | Mcm3 | Phospho | High |

B

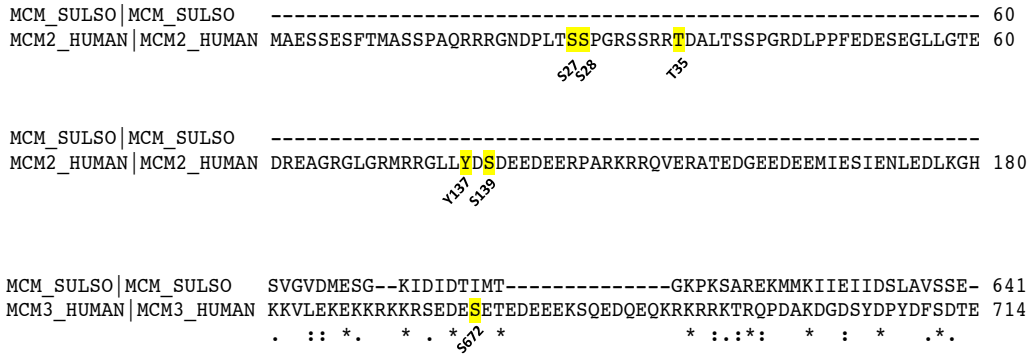



**Figure 31. Phosphorylation in hMCM2-7.** (**A**) Phosphopeptides identified for the hMCM2-7 complex. All peptides are in the high-medium confidence range and described in UniProt or PhosphoSite databases. (**B**) Alignment of the human Mcm2 and Mcm3 with the *Sulfolobus sulfactaricus* Mcm, which its crystal structure is known. None of the phosphorylated residues in the human Mcm subunits are conserved in the SsoMCM.

We identified specific phosphorylations needed for the assembly of the CMG complex (**Figure 31 A**). These phosphorylations were specific for the CMG complex in replication and were not identified for the assembly of the complex in mitosis. We corroborate also that these phosphorylations were not present in the purified hMCM2-7 complex, indicating that these peptides were not phosphorylated during the overexpression of the complex in insect cells.

All phosphorylations detected for hMcm2 are located at on its N-terminal domain. However, we could not localize these phosphorylated residues in the crystal structure of the *Sulfolobus sulfacataricus* MCM complex, as they are not conserved in archaea (**Figure 31 B**).

The detected phosphorilaton for the hMcm3 is located at the C-terminal domain. Although this domain presents certain homology with the carboxy-terminal of SsoMCM, the S672 residue of the hMcm3 is not conserved in archaea (**Figure 31 B**).

# 5. Discussion

## 5.1. 3D structure of the human MCM2-7 complex

hMCM2-7 has been purified in its active form following the protocol described in Material and Methods. As we have already stressed, the presence of 10% glycerol in the buffer was mandatory for the stability of the complex. Loading the protein sample to gel filtration column, the elution volume of the complex corresponds to the molecular weight of a globular protein with a size of 550 kDa. However we couldn't confirm this value by native mass spectrometry due to the disassembling of the complex when glycerol was removed in a buffer exchange process required by this technique. Nevertheless the presence of the six proteins of the hMCM2-7 complex was confirmed by Western-blot and mass spectrometry analysis. EM analysis revealed that the hMCM2-7 complex without nucleotide presents high conformational heterogeneity that is less pronounced in the presence of ADP. Addition of ATPγS further stabilizes the complex while DNA binding introduces important conformational changes by reducing the size of the Mcm2/5 gate as it has been also demonstrated for other eukaryotic MCM2-7 complexes (Bochman and Schawacha, 2008; Costa et al., 2011).

### 5.1.1. Structure of the hMCM2-7 – ATPγS complex

The 3D structure obtained through a single particle negative-stain electron microscopy of the human MCM2-7 complex with ATPγS showed a hexameric toroidal complex with a central cavity similar to the archaeal MCM complexes (Pape et al., 2003) and eukaryotic MCM2-7 complexes (Bochman & Schwacha 2008; Remus et al., 2009; Costa et al., 2011). The side views show two high density parallel bands connected through fine bridges of low protein density. The final 3D structure of the hMCM2-7–ATPγS complex was generated at 24 Å resolution, without applying C6 symmetry. The approximate dimensions of the hMCM2-7 complex are 100 Å high, 130 Å large and 130 Å wide (Figure X). Different starting models were used at the beginning of the refinement. The starting models used were a Gaussian sphere, a noise and a model generated using a common lines and applying C6 symmetry. After several refinement cycles, all reference volumes result in a similar final volume. During the image processing we searched for different conformations of the complex. The hMCM2-7 in presence of ATPγS was present in only one conformation, in contrast to the Drosophila MCM2-7 complex, that in presence of ADP-$BeF_3$ had two different conformations (Costa et al., 2011).

The hMCM2-7–ATPγS is composed of six different subunits forming a ring-like shape with a central cavity and six side channels. The amino-terminal part of the hMCM2-7 complex is located on the bottom/top part of the structure and shows a closed ring with a 30 Å diameter whereas the carboxy-terminal part of the complex, located on the opposite part of the structure, is slightly bigger and presents a discontinuity in the ring. We assume that discontinuity is located between Mcm2 and Mcm5 subunits, as it was described for other eukaryotic MCM complexes (Ref Costa et all). This Mcm2 and Mcm5 gate might be responsible for turning on the helicase activity of the complex as well as for facilitating the loading of the complex onto the DNA (Bochman and Schwacha 2008, Remus et al., 2009; Costa et al., 2011). In our model, the carboxy terminal ends of Mcm2 and Mcm5, in presence of ATPγS, are separated aproximately 15 Å. In this structure the central channel is large enough to fit dsDNA. The six side channels have a diameter of 20 Å, large enough to allow ssDNA through. These side channels located between pairs of subunits are proposed is some models as a possible route to extrude the single stranded DNA during unwinding.

### 5.1.2. Structure of the hMCM2-7 – ATPγS – overhangDNA complex

To obtain the 3D volume of the hMCM2-7 complex bound to DNA (see Material and Methods) we also performed single particle negative stain electron microscopy. The top/bottom views of the reference-free 2D averages show more defined subunits around the central cavity, when compared to the hMCM2-7–ATP $\gamma$ S bound form. On the other hand the side views of the reference-free 2D averages of the complex did not present significant differences comparing to the previous volume. Nevertheless, the 3D structure revealed a conformational change upon DNA binding. The size of the complex bound to overhangDNA is 110 Å high, 125 Å large and 130 Å wide, therefore the hMCM2-7 complex becomes longer and tighter due to the interaction with the DNA. To obtain the final 3D structure we followed the same strategy as for the complex lacking DNA. Also here, only one conformation was present in our data set. The hMCM2-7 complex bound to overhangDNA is an asymmetric structure, with a closed amino-terminal ring around a cavity with 23 Å diameter, and a carboxy-terminal part presents an opened-ring structure. This aperture between the C-terminal part of the subunits Mcm2 and Mcm5 is much smaller than for the volume lacking DNA, having a distance of approximately 3 Å. For performing the hMCM2-7 helicase activity, the Mcm2 and Mcm5 subunits of the complex must be in contact to each other as the MCM2-7 mutants that avoid the binding of Mcm2 and Mcm5 lack helicase

activity (Schwacha and Bell, 2001; Bochman and Scwacha, 2008). In the CMG complex, GINS and Cdc45 are situated around this gate, probably to ensure its closure enhancing the helicase activity of the MCM2-7 (Moyer et al., 2006; Ilves et al., 2010). In spite the conformational changes of the complex bound to DNA, the side view of the complex, still present the six lateral channels, with a diameter large enough to fit ssDNA.

### 5.1.3. Comparison of the different hMCM2-7 structures

Comparing the hMCM2-7 complexes in presence and in absence of DNA, we could observe that DNA binding produces a movement of the central protein loops that connect the carboxy-terminal and amino-terminal parts of the complex. In absence of DNA, these loops are orientated towards the central cavity, however when the complex is loaded onto the overhangDNA, these loops stay parallel to the central channel, making the complex 10 Å larger. This movement makes the complex even more rigid. The hMCM2-7 when bound to overhangDNA has a length of approximately 110 Å, large enough to cover 34 bp of ssDNA. Due to the negative staining technique, the density corresponding to the DNA is not visible in the volume of the hMCM2-7. However, we could see some density coming out from the complex when we processed the cryo-EM data of the hMCM2-7 bound to overhangDNA. In spite of the conformational changes observed, the side channels keep apparently the same size upon DNA binding to hMCM2-7 complex (**Figure 32**).

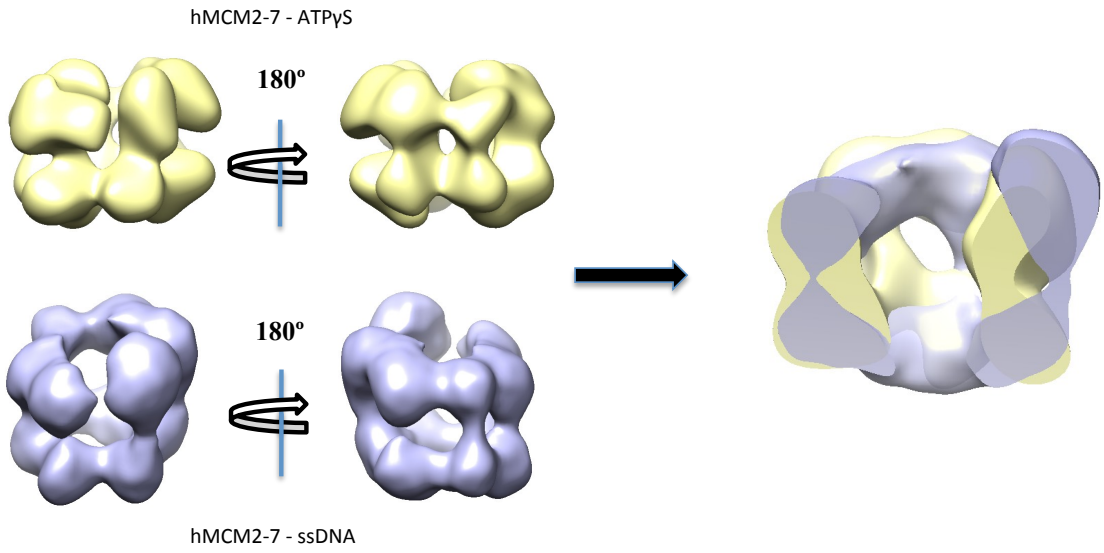



**Figure 32. Comparison of the hMCM2-7 in presence/absence of DNA**. The comparison was done using the lateral views of the complex bound to ATPγS (yellow) and the complex bound to overhangDNA in presence of ATPγS (blue).

### 5.1.4. The Mcm2/Mcm5 "gate"

The gap existing between the Mcm2 and Mcm5 subunits is visible in our 3D models. However we could observe some conformational differences depending on the presence of ATPγS. It is known that Mcm2 and Mcm5 have little physical interaction. This discontinuity has been also observed in the structure of the Drosophila MCM2-7 complex (Costa et al., 2011) and in yeast MCM2-7 (Sun J. et al., 2013). We have tried to solve the 3D structure of the hMCM2-7 in absence of nucleotide. However the flexibility that the complex adopts without any nucleotide enabled us to build the 3D volume of the apo hMCM2-7. After incubation of the complex with the non-hydrolysable ATPγS, the hMCM2-7 was less flexible and the volume we obtained elucidates its opened-ring structure. In this volume the separation at the C-terminal end of the Mcm2 and Mcm5 subunits was around 15 Å. Mutations that avoid the interaction between the Mcm2 and Mcm5 by blocking its binding to the ATPase active site (Mcm2 RA) or by avoiding its physical interaction (Mcm5 KA), are known to block its helicase activity (Bochman and Schwacha, 2007, Bochman and Schwacha, 2008). Thus, the interaction between these subunits is mandatory for the helicase activity of the complex. The Mcm2-5 gate also would facilitate the loading of the complex onto the dsDNA. Cdt1, the protein responsible for loading the MCM2-7 complex to the origins of replication, keeps the MCM2-7 complex in an opened conformation and thus inactive by interacting primarily with Mcm2 but also with Mcm5 and Mcm6. The hMCM2-7 on the overhangDNA in presence of ATPγS has the gap between Mcm2 and Mcm5 almost closed, with a distance of around 3 Å (**Figure 33**). This gap disappears when the complex is visualized with the threshold corresponding to the 100% mass of the hMCM2-7 complex, however for the 95% of the mass, the gate between Mcm2 and Mcm5 becomes visible.

The MCM2-7 complex acts as the replicative helicase *in vivo* whose activity is enhanced through its interaction with GINS and Cdc45 (Ilves I et al., 2010), forming the CMG complex. Structural studies have demonstrated that GINS and Cdc45 would close the MCM2-7 ring as they bridge the gap existing between Mcm2 and Mcm5, in presence of ADP-$BeF_3$ (Costa et al., 2011).

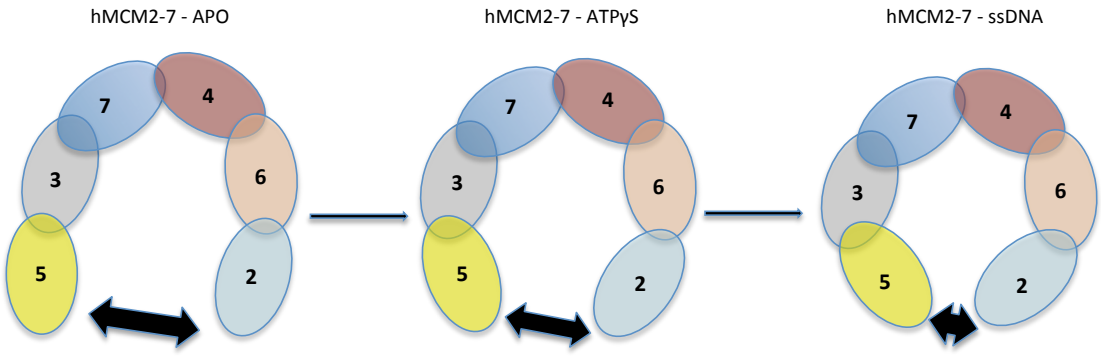



**Figure 33. Representation of the conformational changes of the hMCM2-7 upon ATPγS and DNA binding.**

## 5.2. Helicase activity of the hMCM2-7 complex

The helicase assay showed that human MCM2-7 has *in vitro* 3' to 5' helicase activity, and this helicase activity is dependent on ATP hydrolysis. In addition, ATP binding increases the affinity of the complex for the ssDNA, rather than ATP hydrolysis (Bochman and Schwacha, 2007). The affinity of the MCM2-7 for ssDNA is 100-fold greater than for dsDNA binding, and most importantly, the binding to ssDNA does not require any other protein cofactors as it is the case for the MCM2-7 loading onto dsDNA. The *in vitro* helicase activity of *S. cerevisiae* MCM2-7 complex is anion-dependent, since glutamate enhances its ability to bind ssDNA (Bochman and Schwacha, 2008), probably because it mimics some of the effects of ATP. This anion-dependence occurs also in the case of the hMCM2-7 complex. The glutamate increases the helicase activity of the hMCM2-7 complex whereas the chloride anion competes with the glutamate decreasing significantly its helicase activity, as occurs in yeast MCM2-7 complex (Bochman and Schwacha, 2008).

## 5.3. The CMG complex, as the replicative helicase complex

This CMG complex, composed by Cdc45, MCM2-7 and GINS (Aparicio et al., 2006) is described as the truth replicative helicase in eukaryotes. Although the MCM2-7 complex has helicase activity *in vitro* by itself, Cdc45 and GINS are thought to act as cofactors that increase the helicase activity due to an allosteric change in the MCM2-7 complex. This association of GINS and Cdc45 help to the correct coordination of the six MCM2-7 subunits (Ilves et al., 2010).

In the first steps of DNA replication the MCM2-7 complex forms a double-hexamer on dsDNA and together with ORC complex, Cdc6 and Cdt1 forms the pre-Replicative complex (Evrin et al., 2009; Remus et al., 2009). Although the MCM2-7 appears as double-hexamer at the origins of replication, when the CMG complex is formed, the MCM2-7 is described as a single-hexamer (Moyer et al., 2006). This fact lead to the hypothesis that the MCM2-7 is loaded as a double-hexamer at the origins of replications, but when gets active by its association with Cdc45 and GINS thanks to the kinases CDK and DDK, the two MCM2-7 complexes would form two replication forks that would move along ssDNA in opposite directions (Yardimci et al., 2010).

The MCM2-7 forms the double-hexamer through its N-terminal domains (Remus et al., 2009). Cdc45 and GINS might be responsible for bracking this MCM2-7 double-hexamer. It is known, by structural studies in the Drosophila CMG complex, that the binding of these two cofactors to the MCM2-7 bridges the gap between Mcm2 and Mcm5 located at the C-terminal part of the complex (Costa et al., 2011). However, in the CMG apo structure, GINS keeps extensive contacts with the N-terminal domain of Mcm3 and Mcm5 through its subunits Psf2 and Psf3. Also Cdc45 appears, in this model, to share a large interaction surface with the N-terminal domain of Mcm2. These facts make possible that GINS and Cdc45 would disassemble the double-hexamer and activate the replication fork. However, further studies must be carried on to verify this hypothesis and to understand this mechanism.

For the assembling of the CMG complex, some post-translational modifications on MCM subunits are mandatory. We have proved that incubating purified MCM2-7, Cdc45 and GINS, the CMG complex cannot be assembled *in vitro*. The published works about CMG complex, obtain the complex by isolation the endogenous CMG or by baculovirus co-expression of the eleven subunits of the complex (Moyer et al., 2006, Ilves et al., 2009; Costa et al., 2011), in such a way that the complex assembly occurs in the cell. We have obtained *in vitro* the CMG complex by mimicking the environment required for the CMG reconstitution in the cell. Using interphase G1/S synchronized HeLa cell extract and purified hMCM2-7, GINS and Cdc45, we could reconstitute *in vitro* the CMG complex in presence of the overhangDNA. The existence of all CMG components were confirmed by Western blots against the individual proteins and by mass-spectrometry analysis. Furthermore, in the mass-

spectometry data we could detect most of the proteins described as replisome-associated cofactors in the vicinity of nascent DNA molecules (Lopez-Contreras et al., 2013). In this assay, we found that the CMG complex remains stable after a pull-down experiment using the His-tag at the N-terminal of the hMcm4.

### 5.3.1. Essential phosphorylations for the CMG assembly

In the reconstituted CMG complex, we could detect, by trypsin-digested mass spectrometry, specific phosphopeptides of the hMCM2-7 complex present only in the DNA replication phase. The phosphorylated residues were not present when the complex was assembled in presence of mitotic cell extract neither in the purified hMCM2-7 complex over-expressed in the insect cells. Most of the high-confidence phosphopeptides found in this assay, belong to the amino-terminal domain of Mcm2 while additional phosphopeptide was found at the C-terminal domain of Mcm3. All these phosphopeptiedes are located away from the highly conserved AAA+ ATPase region (**Figure 34 A-B**).

There are very few crystal structures of the Mcm proteins, and all of them belong to archaeal MCMs. We aligned the sequences of the human Mcm2 and Mcm3 with their homologs in other eukaryotes and sequence of *S. solfataricus* MCM whose crystal-structure is known. We found that none of the phosphorylated residues of the human Mcm2 and Mcm3 are present in the SsoMcm, so we couldn't localize these residues in the crystal structure. However, it is known by the electron microscopy structure of the *Drosophila* CMG complex that GINS interact with the N-terminal domain of MCM2-7, mainly with Mcm5 and Mcm3. Cdc45 has also a large contact surface with the N- terminal of Mcm2 (Costa et al., 2011).

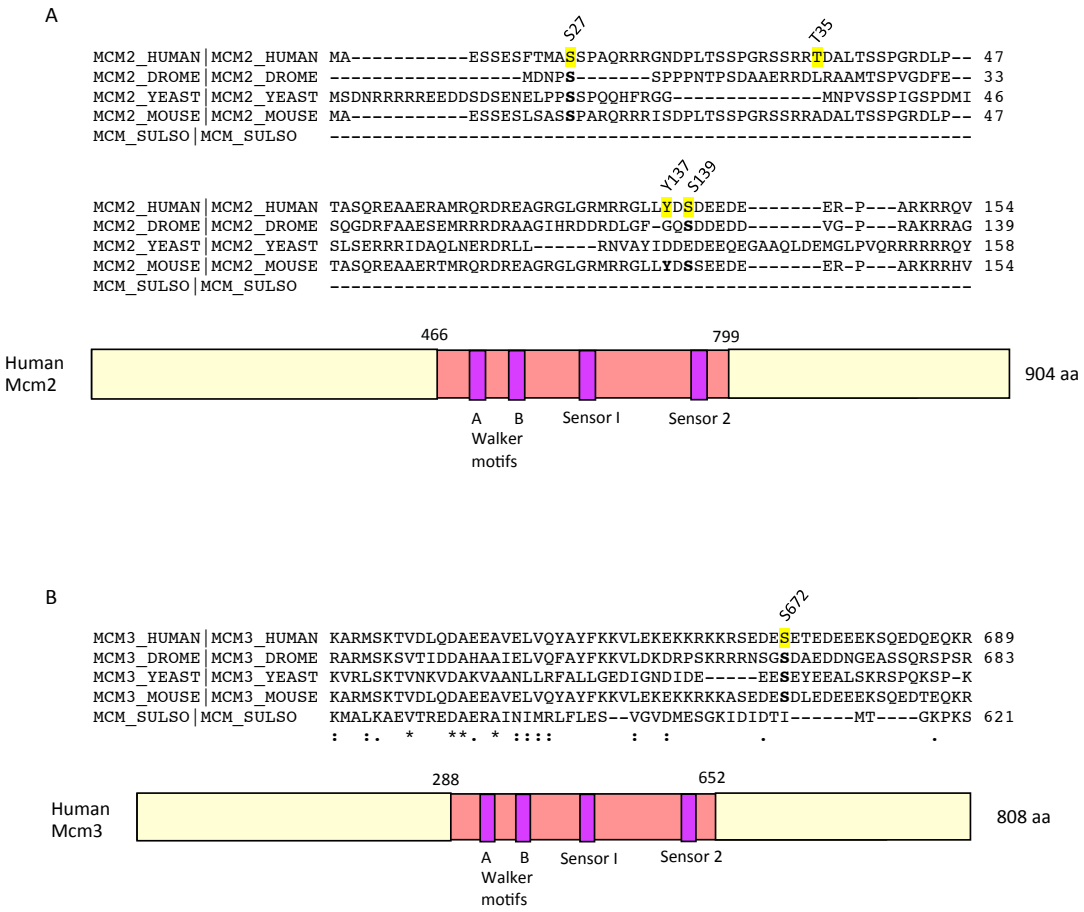



**Figure 34. Alignment of eukaryotic Mcm2 and Mcm3 and SsoMcm and schema of the hMcm2 and hMcm3**. (A) Alignment of hMcm2 with other eukaryotic Mcm2 and SsoMCM. Phosphorylated residues of hMcm2 (yellow) and their positions are indicated. The conserved residues are labeled with bold letters. At the lower part, the hMcm2 is represented with the common motifs indicated. The positions of the AAA+ ATPase domain are indicated. (B) Alignment of hMcm3 with other eukaryotic Mcm3 and SsoMCM. The phosphorylated residue of hMcm3 is colored in yellow and its conserved residues in the other system are marked with bold letters. At the lower part, the hMcm3 is represented with the common motifs indicated. The positions of the AAA+ ATPase domain are indicated.

The Cdc7/Dbf4 kinase is responsible for the *in vitro* and *in vivo* phosphorylation of the residues located in N-terminus of hMcm2. It is known that the Ser27 and Ser139 are direct targets for the Cdc7/Dbf4 kinase during the G1/S phase (Tsuji T. et al., 2006). The sequence alignment of some eukaryotic Mcm2 with the SsoMCM shows that the archaeal mcm protein lacks homology to eukaryotic N-terminus. Probably it is because the N-terminal part of the eukaryotic MCM2-7 complex is involved in the assembly of the CMG complex. The archaeal MCM complexes are full-active helicases without the need of other cofactors as needed in eukaryotes, where the MCM2-7 needs Cdc45 and GINS to develop the full helicase activity. The sequence alignment of the N-terminal domain of Mcm2 also shows that Ser27 is highly conserved in higher and lower eukaryotes, whereas the Ser193 is only conserved from *Drosophila* to mammals. The phosphorylated residues Thr35 and Tyr137 in

hMcm2 and the Ser672 in hMcm3 have been described by proteomic discovery-mode mass spectrometry. The fact that all phosphorylations in hMcm2 that we have detected as a part of the CMG complex, are located at the N-terminal domain, together with the requirement of phosphorylations to assembly the CMG complex, lead us to think that most of these phosphorylated residues are necessary for the assembly of Cdc45, MCM2-7 and GINS in the replication fork. Special attention requires the Ser27 in hMcm2 and Ser672 in hMcm3, residues are conserved among higher and lower eukaryotes.

# 6. Conclusions

1- The human MCM2-7 complex has been overexpressed in Sf21 insect cells and purified in its heterohexameric form.

2- The human MCM2-7 complex is able to unwind dsDNA with 3'-5' direction. Its helicase activity depends on ATP hydrolysis and the presence of certain anions.

3- The human MCM2-7 complex can be loaded onto ssDNA.

4- The three-dimensional reconstruction at low resolution of the human MCM2-7 complex bound to ATPγS shows a six subunits complex with a central cavity and six lateral channels. The N-terminal and C-terminal domains of the complex are asymmetric. There is discontinuity in the ring between the Mcm2 and Mcm5 subunits at the C-terminal domain.

5- The three-dimensional reconstructions of the human MCM2-7 complex show the conformational changes of the complex upon the binding to ADP, ATPγS or ATPγS and DNA.

6- The *in vitro* reconstitution of the CMG complex depends on the presence of ssDNA and the kinase activity of CDKs and DDK that phosphorylate the MCM2-7 complex. Phosphorilations in the N-terminal domain of Mcm2 and Mcm3 are essential for the assembly of the CMG complex.

# 7. Conclusiones

1- El complejo humano MCM2-7 ha sido sobre-expresado en células de insecto Sf21 y purificado en su forma heterohexamérica.

2- El complejo humano MCM2-7 es capaz de desenrollar ADN de doble cadena con dirección 3'-5'. Su actividad helicasa depende de ATP y ciertos aniones.

3- El complejo MCM2-7 puede unirse a cadena sencilla de ADN.

4- La reconstrucción tridimensional a baja resolución del complejo humano MCM2-7 unido a ATPγS muestra un complejo de seis subunidades con una cavidad central y seis canales laterales. Los dominios N-terminal y C-terminal del complejo son asimetricos, presentando una discontinuidad entre las subunidades Mcm2 y Mcm5 en el dominio C-terminal.

5- La reconstrucción tridimensional del complejo humano MCM2-7 muestra los cambios conformacionales que sufre el complejo tras la unión a ADP, ATPγS o ATPγS junto a ADN de cadena sencilla.

6- La reconstitución *in vitro* del complejo CMG depende de la presencia de ADN de cadena sencilla y de la actividad kinasa de CDKs y DDK que fosforilan al complejo MCM2-7. Fosforilaciones en el dominio N-terminal de Mcm2 y Mcm3 son esenciales para el ensamblaje del complejo CMG.

# 8. References

Aparicio, T., E. Guillou, J. Coloma, G. Montoya and J. Mendez (2009). "The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication." Nucleic Acids Res **37**(7): 2087-2095.

Bae, B., Y. H. Chen, A. Costa, S. Onesti, J. S. Brunzelle, Y. Lin, I. K. Cann and S. K. Nair (2009). "Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog." Structure **17**(2): 211-222.

Barry, E. R., A. T. McGeoch, Z. Kelman and S. D. Bell (2007). "Archaeal MCM has separable processivity, substrate choice and helicase domains." Nucleic Acids Res **35**(3): 988-998.

Bauerschmidt, C., S. Pollok, E. Kremmer, H. P. Nasheuer and F. Grosse (2007). "Interactions of human Cdc45 with the Mcm2-7 complex, the GINS complex, and DNA polymerases delta and epsilon during S phase." Genes Cells **12**(6): 745-758.

Bell, S. P. and A. Dutta (2002). "DNA replication in eukaryotic cells." Annu Rev Biochem **71**: 333-374.

Bell, S. P. and B. Stillman (1992). "ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex." Nature **357**(6374): 128-134.

Berger, I., D. J. Fitzgerald and T. J. Richmond (2004). "Baculovirus expression system for heterologous multiprotein complexes." Nat Biotechnol **22**(12): 1583-1587.

Bochman, M. L. and A. Schwacha (2007). "Differences in the single-stranded DNA binding activities of MCM2-7 and MCM467: MCM2 and MCM5 define a slow ATP-dependent step." J Biol Chem **282**(46): 33795-33804.

Bochman, M. L. and A. Schwacha (2008). "The Mcm2-7 complex has in vitro helicase activity." Mol Cell **31**(2): 287-293.

Bochman, M. L. and A. Schwacha (2009). "The Mcm complex: unwinding the mechanism of a replicative helicase." Microbiol Mol Biol Rev **73**(4): 652-683.

Bochman, M. L. and A. Schwacha (2010). "The Saccharomyces cerevisiae Mcm6/2 and Mcm5/3 ATPase active sites contribute to the function of the putative Mcm2-7 'gate'." Nucleic Acids Res **38**(18): 6078-6088.

Boos, D., J. Frigola and J. F. Diffley (2012). "Activation of the replicative DNA helicase: breaking up is hard to do." Curr Opin Cell Biol **24**(3): 423-430.

Boskovic, J., J. Coloma, T. Aparicio, M. Zhou, C. V. Robinson, J. Mendez and G. Montoya (2007). "Molecular architecture of the human GINS complex." EMBO Rep **8**(7): 678-684.

Brewster, A. S. and X. S. Chen (2010). "Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex." Crit Rev Biochem Mol Biol **45**(3): 243-256.

Bruck, I. and D. L. Kaplan (2013). "Cdc45 protein-single-stranded DNA interaction is important for stalling the helicase during replication stress." J Biol Chem **288**(11): 7550-7563.

Chagin, V. O., J. H. Stear and M. C. Cardoso (2010). "Organization of DNA replication." Cold Spring Harb Perspect Biol **2**(4): a000737.

Chang, Y. P., G. Wang, V. Bermudez, J. Hurwitz and X. S. Chen (2007). "Crystal structure of the GINS complex and functional insights into its role in DNA replication." Proc Natl Acad Sci U S A **104**(31): 12685-12690.

Chen, Z., C. Speck, P. Wendel, C. Tang, B. Stillman and H. Li (2008). "The architecture of the DNA replication origin recognition complex in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **105**(30): 10326-10331.

Cho, W. H., Y. J. Lee, S. I. Kong, J. Hurwitz and J. K. Lee (2006). "CDC7 kinase phosphorylates serine residues adjacent to acidic amino acids in the minichromosome maintenance 2 protein." Proc Natl Acad Sci U S A **103**(31): 11521-11526.

Chong, J. P., P. Thommes and J. J. Blow (1996). "The role of MCM/P1 proteins in the licensing of DNA replication." Trends Biochem Sci **21**(3): 102-106.

Chuang, L. C., L. K. Teixeira, J. A. Wohlschlegel, M. Henze, J. R. Yates, J. Mendez and S. I. Reed (2009). "Phosphorylation of Mcm2 by Cdc7 promotes pre-replication complex assembly during cell-cycle re-entry." Mol Cell **35**(2): 206-216.

Clarey, M. G., J. P. Erzberger, P. Grob, A. E. Leschziner, J. M. Berger, E. Nogales and M. Botchan (2006). "Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex." Nat Struct Mol Biol **13**(8): 684-690.

Costa, A., I. Ilves, N. Tamberg, T. Petojevic, E. Nogales, M. R. Botchan and J. M. Berger (2011). "The structural basis for MCM2-7 helicase activation by GINS and Cdc45." Nat Struct Mol Biol **18**(4): 471-477.

Costa, A. and S. Onesti (2008). "The MCM complex: (just) a replicative helicase?" Biochem Soc Trans **36**(Pt 1): 136-140.

Costa, A. and S. Onesti (2009). "Structural biology of MCM helicases." Crit Rev Biochem Mol Biol **44**(5): 326-342.

Costa, A., T. Pape, M. van Heel, P. Brick, A. Patwardhan and S. Onesti (2006). "Structural basis of the Methanothermobacter thermautotrophicus MCM helicase activity." Nucleic Acids Res **34**(20): 5829-5838.

Costa, A., T. Pape, M. van Heel, P. Brick, A. Patwardhan and S. Onesti (2006). "Structural studies of the archaeal MCM complex in different functional states." J Struct Biol **156**(1): 210-219.

Costa, A., G. van Duinen, B. Medagli, J. Chong, N. Sakakibara, Z. Kelman, S. K. Nair, A. Patwardhan and S. Onesti (2008). "Cryo-electron microscopy reveals a novel DNA-binding site on the MCM helicase." EMBO J **27**(16): 2250-2258.

Cox, J., and Mann, M. (2008). “MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.” Nature Biotechnology 26, 1367–1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011). “Andromeda: a peptide search engine integrated into the MaxQuant environment.” J. Proteome Res. 10, 1794–1805.

Davey, M. J. and M. O'Donnell (2003). "Replicative helicase loaders: ring breakers and ring makers." Curr Biol **13**(15): R594-596.

Diffley, J. F. (2011). "Quality control in the initiation of eukaryotic DNA replication." Philos Trans R Soc Lond B Biol Sci **366**(1584): 3545-3553.
Dueber, E. L., J. E. Corn, S. D. Bell and J. M. Berger (2007). "Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex." Science **317**(5842): 1210-1213.

Edwards, M. C., A. V. Tutter, C. Cvetic, C. H. Gilbert, T. A. Prokhorova and J. C. Walter (2002). "MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts." J Biol Chem **277**(36): 33049-33057.

Evrin, C., P. Clarke, J. Zech, R. Lurz, J. Sun, S. Uhle, H. Li, B. Stillman and C. Speck (2009). "A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication." Proc Natl Acad Sci U S A **106**(48): 20240-20245.

Fernandez-Cid, A., A. Riera, S. Tognetti, M. C. Herrera, S. Samel, C. Evrin, C. Winkler, E. Gardenal, S. Uhle and C. Speck (2013). "An ORC/Cdc6/MCM2-7 complex is formed in a multistep reaction to serve as a platform for MCM double-hexamer assembly." Mol Cell **50**(4): 577-588.

Fitzgerald, D. J., P. Berger, C. Schaffitzel, K. Yamada, T. J. Richmond and I. Berger (2006). "Protein complex expression by using multigene baculoviral vectors." Nat Methods **3**(12): 1021-1032.

Fletcher, R. J., B. E. Bishop, R. P. Leon, R. A. Sclafani, C. M. Ogata and X. S. Chen (2003). "The structure and function of MCM from archaeal M. Thermoautotrophicum." Nat Struct Biol **10**(3): 160-167.

Flores-Rozas, H., D. Clark and R. D. Kolodner (2000). "Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex." Nat Genet **26**(3): 375-378.

Forsburg, S. L. (2004). "Eukaryotic MCM proteins: beyond replication initiation." Microbiol Mol Biol Rev **68**(1): 109-131.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L. J. (2013). “STRING v9.1: protein-protein interaction networks, with increased coverage and integration.” Nucleic Acids Res. 41, D808-815. doi: 10.1093/nar/gks1094. Epub 2012 Nov 29.

Fu, Y. V., H. Yardimci, D. T. Long, T. V. Ho, A. Guainazzi, V. P. Bermudez, J. Hurwitz, A. van Oijen, O. D. Scharer and J. C. Walter (2011). "Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase." Cell **146**(6): 931-941.

Gai, D., D. Li, C. V. Finkielstein, R. D. Ott, P. Taneja, E. Fanning and X. S. Chen (2004). "Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen." J Biol Chem **279**(37): 38952-38959.

Gambus, A., R. C. Jones, A. Sanchez-Diaz, M. Kanemaki, F. van Deursen, R. D. Edmondson and K. Labib (2006). "GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks." Nat Cell Biol **8**(4): 358-366.

Gambus, A., G. A. Khoudoli, R. C. Jones and J. J. Blow (2011). "MCM2-7 form double hexamers at licensed origins in Xenopus egg extract." J Biol Chem **286**(13): 11855-11864.

Garcia, V., K. Furuya and A. M. Carr (2005). "Identification and functional analysis of TopBP1 and its homologs." DNA Repair (Amst) **4**(11): 1227-1239.

Hashimoto, Y., F. Puddu and V. Costanzo (2012). "RAD51- and MRE11-dependent reassembly of uncoupled CMG helicase complex at collapsed replication forks." Nat Struct Mol Biol **19**(1): 17-24.

Hombauer, H., C. S. Campbell, C. E. Smith, A. Desai and R. D. Kolodner (2011). "Visualization of eukaryotic DNA mismatch repair reveals distinct recognition and repair intermediates." Cell **147**(5): 1040-1053.

Ilves, I., T. Petojevic, J. J. Pesavento and M. R. Botchan (2010). "Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins." Mol Cell **37**(2): 247-258.

Im, J. S., S. H. Ki, A. Farina, D. S. Jung, J. Hurwitz and J. K. Lee (2009). "Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins." Proc Natl Acad Sci U S A **106**(37): 15628-15632.

Ishimi, Y. (1997). "A DNA helicase activity is associated with an MCM4, -6, and -7 protein complex." J Biol Chem **272**(39): 24508-24513.

Jenkinson, E. R. and J. P. Chong (2006). "Minichromosome maintenance helicase activity is controlled by N- and C-terminal motifs and requires the ATPase domain helix-2 insert." Proc Natl Acad Sci U S A **103**(20): 7613-7618.

Jiricny, J. (2006). "MutLalpha: at the cutting edge of mismatch repair." Cell **126**(2): 239-241.
Johansson, E. and S. A. Macneill (2010). "The eukaryotic replicative DNA polymerases take shape." Trends Biochem Sci **35**(6): 339-347.

Kamimura, Y., H. Masumoto, A. Sugino and H. Araki (1998). "Sld2, which interacts with Dpb11 in Saccharomyces cerevisiae, is required for chromosomal DNA replication." Mol Cell Biol **18**(10): 6102-6109.

Kamimura, Y., Y. S. Tak, A. Sugino and H. Araki (2001). "Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae." EMBO J **20**(8): 2097-2107.

Kanemaki, M., A. Sanchez-Diaz, A. Gambus and K. Labib (2003). "Functional proteomic identification of DNA replication proteins by induced proteolysis in vivo." Nature **423**(6941): 720-724.

Kang, Y. H., A. Farina, V. P. Bermudez, I. Tappin, F. Du, W. C. Galal and J. Hurwitz (2013). "Interaction between human Ctf4 and the Cdc45/Mcm2-7/GINS (CMG) replicative helicase." Proc Natl Acad Sci U S A **110**(49): 19760-19765.

Kanke, M., Y. Kodama, T. S. Takahashi, T. Nakagawa and H. Masukata (2012). "Mcm10 plays an essential role in origin DNA unwinding after loading of the CMG components." EMBO J **31**(9): 2182-2194.

Kaplan, D. L., M. J. Davey and M. O'Donnell (2003). "Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex." J Biol Chem **278**(49): 49171-49182.

Kaplan, D. L. and M. O'Donnell (2004). "Twin DNA pumps of a hexameric helicase provide power to simultaneously melt two duplexes." Mol Cell **15**(3): 453-465.

Kearsey, S. E. and K. Labib (1998). "MCM proteins: evolution, properties, and role in DNA replication." Biochim Biophys Acta **1398**(2): 113-136.

Kelman, Z. and J. Hurwitz (2003). "Structural lessons in DNA replication from the third domain of life." Nat Struct Biol **10**(3): 148-150.

Krastanova, I., V. Sannino, H. Amenitsch, O. Gileadi, F. M. Pisani and S. Onesti (2012). "Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases." J Biol Chem **287**(6): 4121-4128.

Kubota, Y., Y. Takase, Y. Komori, Y. Hashimoto, T. Arata, Y. Kamimura, H. Araki and H. Takisawa (2003). "A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication." Genes Dev **17**(9): 1141-1152.

Kunkel, T. A. and P. M. Burgers (2008). "Dividing the workload at a eukaryotic replication fork." Trends Cell Biol **18**(11): 521-527.

Kurth, I. and M. O'Donnell (2013). "New insights into replisome fluidity during chromosome replication." Trends Biochem Sci **38**(4): 195-203.

Labib, K. (2010). "How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells?" Genes Dev **24**(12): 1208-1219.

Labib, K., J. A. Tercero and J. F. Diffley (2000). "Uninterrupted MCM2-7 function required for DNA replication fork progression." Science **288**(5471): 1643-1647.

Laskey, R. A. and M. A. Madine (2003). "A rotary pumping model for helicase function of MCM proteins at a distance from replication forks." EMBO Rep **4**(1): 26-30.

Lee, J. K. and J. Hurwitz (2001). "Processive DNA helicase activity of the minichromosome maintenance proteins 4, 6, and 7 complex requires forked DNA structures." Proc Natl Acad Sci U S A **98**(1): 54-59.

Lei, M., Y. Kawasaki, M. R. Young, M. Kihara, A. Sugino and B. K. Tye (1997). "Mcm2 is a target of regulation by Cdc7-Dbf4 during the initiation of DNA synthesis." Genes Dev **11**(24): 3365-3374.

Lei, M. and B. K. Tye (2001). "Initiating DNA synthesis: from recruiting to activating the MCM complex." J Cell Sci **114**(Pt 8): 1447-1454.

Leung, C. C., L. Sun, Z. Gong, M. Burkat, R. Edwards, M. Assmus, J. Chen and J. N. Glover (2013). "Structural insights into recognition of MDC1 by TopBP1 in DNA replication checkpoint control." Structure **21**(8): 1450-1459.

Li, C. and J. Jin (2010). "DNA replication licensing control and rereplication prevention." Protein Cell **1**(3): 227-236.

Li, D., R. Zhao, W. Lilyestrom, D. Gai, R. Zhang, J. A. DeCaprio, E. Fanning, A. Jochimiak, G. Szakonyi and X. S. Chen (2003). "Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen." Nature **423**(6939): 512-518.

Lipps, G., A. O. Weinzierl, G. von Scheven, C. Buchen and P. Cramer (2004). "Structure of a bifunctional DNA primase-polymerase." Nat Struct Mol Biol **11**(2): 157-162.

Liu, C., R. Wu, B. Zhou, J. Wang, Z. Wei, B. K. Tye, C. Liang and G. Zhu (2012). "Structural insights into the Cdt1-mediated MCM2-7 chromatin loading." Nucleic Acids Res **40**(7): 3208-3217.

Liu, J., C. L. Smith, D. DeRyckere, K. DeAngelis, G. S. Martin and J. M. Berger (2000). "Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control." Mol Cell **6**(3): 637-648.

Liu, W., B. Pucci, M. Rossi, F. M. Pisani and R. Ladenstein (2008). "Structural analysis of the Sulfolobus solfataricus MCM protein N-terminal domain." Nucleic Acids Res **36**(10): 3235-3243.

Lopes, M., C. Cotta-Ramusino, A. Pellicioli, G. Liberi, P. Plevani, M. Muzi-Falconi, C. S. Newlon and M. Foiani (2001). "The DNA replication checkpoint response stabilizes stalled replication forks." Nature **412**(6846): 557-561.

Lopez-Contreras, A. J., I. Ruppen, M. Nieto-Soler, M. Murga, S. Rodriguez-Acebes, S. Remeseiro, S. Rodrigo-Perez, A. M. Rojas, J. Mendez, J. Munoz and O. Fernandez-Capetillo (2013). "A proteomic characterization of factors enriched at nascent DNA molecules." Cell Rep **3**(4): 1105-1116.

Ludtke, S. J., P. R. Baldwin and W. Chiu (1999). "EMAN: semiautomated software for high-resolution single-particle reconstructions." J Struct Biol **128**(1): 82-97.

Machida, Y. J., J. L. Hamlin and A. Dutta (2005). "Right place, right time, and only once: replication initiation in metazoans." Cell **123**(1): 13-24.

MacNeill, S. A. (2010). "Structure and function of the GINS complex, a key component of the eukaryotic replisome." Biochem J **425**(3): 489-500.

Maga, G., G. Villani, V. Tillement, M. Stucki, G. A. Locatelli, I. Frouin, S. Spadari and U. Hubscher (2001). "Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1." Proc Natl Acad Sci U S A **98**(25): 14298-14303.

Maine, G. T., P. Sinha and B. K. Tye (1984). "Mutants of S. cerevisiae defective in the maintenance of minichromosomes." Genetics **106**(3): 365-385.

Masai, H., C. Taniyama, K. Ogino, E. Matsui, N. Kakusho, S. Matsumoto, J. M. Kim, A. Ishii, T. Tanaka, T. Kobayashi, K. Tamai, K. Ohtani and K. Arai (2006). "Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin." J Biol Chem **281**(51): 39249-39261.
McGeoch, A. T. and S. D. Bell (2005). "Eukaryotic/archaeal primase and MCM proteins encoded in a bacteriophage genome." Cell **120**(2): 167-168.

Mendez, J. and B. Stillman (2003). "Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins." Bioessays **25**(12): 1158-1167.

Meselson, M. and F. W. Stahl (1958). "The replication of DNA." Cold Spring Harb Symp Quant Biol **23**: 9-12.

Mindell, J. A. and N. Grigorieff (2003). "Accurate determination of local defocus and specimen tilt in electron microscopy." J Struct Biol **142**(3): 334-347.

Moir, D., S. E. Stewart, B. C. Osmond and D. Botstein (1982). "Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies." Genetics **100**(4): 547-563.

Montagnoli, A., B. Valsasina, V. Croci, M. Menichincheri, S. Rainoldi, V. Marchesi, M. Tibolla, P. Tenca, D. Brotherton, C. Albanese, V. Patton, R. Alzani, A. Ciavolella, F. Sola, A. Molinari, D. Volpi, N. Avanzi, F. Fiorentini, M. Cattoni, S. Healy, D. Ballinari, E. Pesenti, A. Isacchi, J. Moll, A. Bensimon, E. Vanotti and C. Santocanale (2008). "A Cdc7 kinase inhibitor restricts initiation of DNA replication and has antitumor activity." Nat Chem Biol **4**(6): 357-365.

Moreau, M. J., A. T. McGeoch, A. R. Lowe, L. S. Itzhaki and S. D. Bell (2007). "ATPase site architecture and helicase mechanism of an archaeal MCM." Mol Cell **28**(2): 304-314.

Mossi, R., R. C. Keller, E. Ferrari and U. Hubscher (2000). "DNA polymerase switching: II. Replication factor C abrogates primer synthesis by DNA polymerase alpha at a critical length." J Mol Biol **295**(4): 803-814.

Moyer, S. E., P. W. Lewis and M. R. Botchan (2006). "Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase." Proc Natl Acad Sci U S A **103**(27): 10236-10241.

Nishiyama, A., L. Frappier and M. Mechali (2011). "MCM-BP regulates unloading of the MCM2-7 helicase in late S phase." Genes Dev **25**(2): 165-175.

Okorokov, A. L., A. Waugh, J. Hodgkinson, A. Murthy, H. K. Hong, E. Leo, M. B. Sherman, K. Stoeber, E. V. Orlova and G. H. Williams (2007). "Hexameric ring structure of human MCM10 DNA replication factor." EMBO Rep **8**(10): 925-930.

Onesti, S. and S. A. MacNeill (2013). "Structure and evolutionary origins of the CMG complex." Chromosoma **122**(1-2): 47-53.

Panuska, J. R. and D. A. Goldthwait (1980). "A DNA-dependent ATPase from T4-infected Escherichia coli. Purification and properties of a 63,000-dalton enzyme and its conversion to a 22,000-dalton form." J Biol Chem **255**(11): 5208-5214.

Pasero, P., K. Shimada and B. P. Duncker (2003). "Multiple roles of replication forks in S phase checkpoints: sensors, effectors and targets." Cell Cycle **2**(6): 568-572.

Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. (1999). “Probability-based protein identification by searching sequence databases using mass spectrometry data.” Electrophoresis 20, 3551-3567.

Pollok, S. and F. Grosse (2007). "Cdc45 degradation during differentiation and apoptosis." Biochem Biophys Res Commun **362**(4): 910-915.

Pucci, B., M. De Felice, M. Rossi, S. Onesti and F. M. Pisani (2004). "Amino acids of the Sulfolobus solfataricus mini-chromosome maintenance-like DNA helicase involved in DNA binding/remodeling." J Biol Chem **279**(47): 49222-49228.

Randell, J. C., J. L. Bowers, H. K. Rodriguez and S. P. Bell (2006). "Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase." Mol Cell **21**(1): 29-39.

Remus, D. and J. F. Diffley (2009). "Eukaryotic DNA replication control: lock and load, then fire." Curr Opin Cell Biol **21**(6): 771-777.

Rothenberg, E., M. A. Trakselis, S. D. Bell and T. Ha (2007). "MCM forked substrate specificity involves dynamic interaction with the 5'-tail." J Biol Chem **282**(47): 34229-34234.

Saikrishnan, K., S. P. Griffiths, N. Cook, R. Court and D. B. Wigley (2008). "DNA binding to RecD: role of the 1B domain in SF1B helicase activity." EMBO J **27**(16): 2222-2229.

Sakakibara, N., L. M. Kelman and Z. Kelman (2009). "How is the archaeal MCM helicase assembled at the origin? Possible mechanisms." Biochem Soc Trans **37**(Pt 1): 7-11.

Sakakibara, N., L. M. Kelman and Z. Kelman (2009). "Unwinding the structure and function of the archaeal MCM helicase." Mol Microbiol **72**(2): 286-296.

Samson, R. Y. and S. D. Bell (2013). "MCM loading--an open-and-shut case?" Mol Cell **50**(4): 457-458.

Sanchez-Berrondo, J., P. Mesa, A. Ibarra, M. I. Martinez-Jimenez, L. Blanco, J. Mendez, J. Boskovic and G. Montoya (2012). "Molecular architecture of a multifunctional MCM complex." Nucleic Acids Res **40**(3): 1366-1380.

Sato, M., T. Gotow, Z. You, Y. Komamura-Kohno, Y. Uchiyama, N. Yabuta, H. Nojima and Y. Ishimi (2000). "Electron microscopic observation and single-stranded DNA binding activity of the Mcm4,6,7 complex." J Mol Biol **300**(3): 421-431.

Scheres, S. H., R. Marabini, S. Lanzavecchia, F. Cantele, T. Rutten, S. D. Fuller, J. M. Carazo, R. M. Burnett and C. San Martin (2005). "Classification of single-projection reconstructions for cryo-electron microscopy data of icosahedral viruses." J Struct Biol **151**(1): 79-91.

Schwacha, A. and S. P. Bell (2001). "Interactions between two catalytically distinct MCM subgroups are essential for coordinated ATP hydrolysis and DNA replication." Mol Cell **8**(5): 1093-1104.

Sclafani, R. A. and T. M. Holzen (2007). "Cell cycle regulation of DNA replication." Annu Rev Genet **41**: 237-280.

Seki, T., M. Akita, Y. Kamimura, S. Muramatsu, H. Araki and A. Sugino (2006). "GINS is a DNA polymerase epsilon accessory factor during chromosomal DNA replication in budding yeast." J Biol Chem **281**(30): 21422-21432.

Sheu, Y. J. and B. Stillman (2006). "Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression." Mol Cell **24**(1): 101-113.

Sheu, Y. J. and B. Stillman (2010). "The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4." Nature **463**(7277): 113-117.

Shin, J. H., Y. Jiang, B. Grabowski, J. Hurwitz and Z. Kelman (2003). "Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases." J Biol Chem **278**(49): 49053-49062.

Shin, J. H. and Z. Kelman (2006). "The replicative helicases of bacteria, archaea, and eukarya can unwind RNA-DNA hybrid substrates." J Biol Chem **281**(37): 26914-26921.

Singleton, M. R., M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski and D. B. Wigley (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-193.

Sorzano, C. O., R. Marabini, J. Velazquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo and A. Pascual-Montano (2004). "XMIPP: a new generation of an open-source image processing package for electron microscopy." J Struct Biol **148**(2): 194-204.

Sun, J., C. Evrin, S. A. Samel, A. Fernandez-Cid, A. Riera, H. Kawakami, B. Stillman, C. Speck and H. Li (2013). "Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA." Nat Struct Mol Biol **20**(8): 944-951.

Szambowska, A., I. Tessmer, P. Kursula, C. Usskilat, P. Prus, H. Pospiech and F. Grosse (2014). "DNA binding properties of human Cdc45 suggest a function as molecular wedge for DNA unwinding." Nucleic Acids Res **42**(4): 2308-2319.

Takahashi, T. S., D. B. Wigley and J. C. Walter (2005). "Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase." Trends Biochem Sci **30**(8): 437-444.

Takayama, Y., Y. Kamimura, M. Okawa, S. Muramatsu, A. Sugino and H. Araki (2003). "GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast." Genes Dev **17**(9): 1153-1165.

Takeda, D. Y. and A. Dutta (2005). "DNA replication and progression through S phase." Oncogene **24**(17): 2827-2843.

Takeda, D. Y., Y. Shibata, J. D. Parvin and A. Dutta (2005). "Recruitment of ORC or CDC6 to DNA is sufficient to create an artificial origin of replication in mammalian cells." Genes Dev **19**(23): 2827-2836.

Tanaka, S. and H. Araki (2013). "Helicase activation and establishment of replication forks at chromosomal origins of replication." Cold Spring Harb Perspect Biol **5**(12): a010371.

Tanaka, S., Y. S. Tak and H. Araki (2007). "The role of CDK in the initiation step of DNA replication in eukaryotes." Cell Div **2**: 16.

Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees and S. J. Ludtke (2007). "EMAN2: an extensible image processing suite for electron microscopy." J Struct Biol **157**(1): 38-46.

Tenca, P., D. Brotherton, A. Montagnoli, S. Rainoldi, C. Albanese and C. Santocanale (2007). "Cdc7 is an active kinase in human cancer cells undergoing replication stress." J Biol Chem **282**(1): 208-215.

Treisman, R. and G. Ammerer (1992). "The SRF and MCM1 transcription factors." Curr Opin Genet Dev **2**(2): 221-226.

Tsuji, T., S. B. Ficarro and W. Jiang (2006). "Essential role of phosphorylation of MCM2 by Cdc7/Dbf4 in the initiation of DNA replication in mammalian cells." Mol Biol Cell **17**(10): 4459-4472.

Tuteja, N. and R. Tuteja (2004). "Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery." Eur J Biochem **271**(10): 1835-1848.

Tuteja, N. and R. Tuteja (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function." Eur J Biochem **271**(10): 1849-1863.

Tye, B. K. (1999). "MCM proteins in DNA replication." Annu Rev Biochem **68**: 649-686.
Waga, S. and B. Stillman (1998). "The DNA replication fork in eukaryotic cells." Annu Rev Biochem **67**: 721-751.

Warren, E. M., H. Huang, E. Fanning, W. J. Chazin and B. F. Eichman (2009). "Physical interactions between Mcm10, DNA, and DNA polymerase alpha." J Biol Chem **284**(36): 24662-24672.

Watson, J. D. and F. H. Crick (1953). "The structure of DNA." Cold Spring Harb Symp Quant Biol **18**: 123-131.

Weinreich, M. and B. Stillman (1999). "Cdc7p-Dbf4p kinase binds to chromatin during S phase and is regulated by both the APC and the RAD53 checkpoint pathway." EMBO J **18**(19): 5334-5346.

Wisniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). “Universal sample preparation method for proteome analysis.” Nature Methods 6, 359–362

Woese, C. R. and G. E. Fox (1977). "The concept of cellular evolution." J Mol Evol **10**(1): 1-6.

Xu, M., Y. P. Chang and X. S. Chen (2013). "Expression, purification and biochemical characterization of Schizosaccharomyces pombe Mcm4, 6 and 7." BMC Biochem **14**: 5.

Yardimci, H. and J. C. Walter (2014). "Prereplication-complex formation: a molecular double take?" Nat Struct Mol Biol **21**(1): 20-25.

You, Z., Y. Komamura and Y. Ishimi (1999). "Biochemical analysis of the intrinsic Mcm4-Mcm6-mcm7 DNA helicase activity." Mol Cell Biol **19**(12): 8003-8015.

Zegerman, P. and J. F. Diffley (2007). "Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast." Nature **445**(7125): 281-285.